

THESIS / THÈSE

LICENCE EN BIOLOGIE

Le promoteur du gène reelin de la souris

Gillet, Jean-Pierre

Award date:
2000

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**FACULTÉS UNIVERSITAIRES NOTRE-DAME DE LA PAIX
NAMUR**

Faculté des Sciences

Le promoteur du gène *reelin* de la souris

**Mémoire présenté pour l'obtention du grade de
licencié en Sciences biologiques**

Jean-Pierre GILLET

Juin 2000

**Facultés Universitaires Notre-Dame de la Paix
FACULTE DES SCIENCES**

Secrétariat du Département de Biologie

Rue de Bruxelles 61 - 5000 NAMUR

Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20

E-mail: joelle.jonet@fundp.ac.be - <http://www.fundp.ac.be/fundp.html>

Le promoteur du gène *reelin* de la souris

GILLET Jean-Pierre

Résumé

Le gène *reelin* détient un rôle central dans le développement cérébral embryonnaire et est exprimé de manière particulièrement intense par les neurones de Cajal-Retzius de la zone marginale du cortex. Alors que la structure génomique du gène et de nombreuses informations concernant la voie de signalisation de la protéine *reelin* sont connues, on dispose de très peu d'informations sur la régulation transcriptionnelle de *reelin*. C'est pourquoi nous avons commencé une étude du promoteur du gène *reelin* chez la souris.

Nous avons montré que, chez la souris comme chez l'homme, le gène *reelin* est dépourvu de boîte TATA et des autres éléments classiques de régulation transcriptionnelle, mais contient par contre un îlot CpG de 1.4kb couvrant une partie de la région promotrice et environ 500pb de l'exon 1. Nous avons montré que le pattern de méthylation de cet îlot est invariant à divers stades du développement et dans divers tissus, de sorte que la méthylation n'intervient probablement pas de manière cruciale dans la régulation transcriptionnelle.

Par des programmes de prédiction, nous avons mis en évidence la présence de nombreux sites potentiels de fixation pour des facteurs transcriptionnels, parmi lesquels des facteurs comme Pax5, spécifiques du tissu nerveux. Ces résultats peuvent servir de préalable à des analyses biochimiques des sites de fixation.

Une séquence de 1.7 kb contenant le promoteur minimal de souris a été ensuite étudiée par transfections transitoires de cinq constructions réalisées par délétions progressives dans des cellules P19. Les résultats préliminaires suggèrent qu'il existe deux sites potentiels de fixation de facteurs stimulants séparés par un site potentiel de fixation d'un facteur inhibiteur. La comparaison avec le promoteur humain montre une certaine similitude puisque ce dernier contient aussi deux régions activatrices flanquant une région inhibitrice.

Afin de disposer de l'ensemble de la région génomique contenant les sites de régulation, nous avons isolé un clone PAC de souris qui contient 20kb de DNA génomique en amont du site d'initiation, ainsi que les premiers exons de *reelin* et surtout l'ensemble du premier intron dont la taille est d'environ 70-80 kb. Ce clone devrait s'avérer très utile pour cibler les régions régulatrices par analyse transgénique in vivo.

Mémoire de licence en Sciences Biologiques

Juin 2000

Promoteur: A. Goffinet, Laboratoire de Neurobiologie. Médecine.

Quatre années d'études viennent de s'écouler et citer toutes les personnes qui m'ont soutenue d'une manière ou d'une autre durant ce « long » périple serait une entreprise fastidieuse.

Cependant, au terme de ce travail, je ne peux m'empêcher de remercier ...

Monsieur le professeur Goffinet, merci à vous de m'avoir accueilli dans votre laboratoire,

Madame et messieurs les professeurs Raes et Van Cutsem, Pascal Mertens et Emmanuel Minet, mes lecteurs. Merci à vous de consacrer une partie de votre temps sur ce travail, et vous souhaite une bonne lecture,

Isabelle Bar, un très grand merci à toi pour ta grande disponibilité, ta patience et tes nombreux conseils judicieux tout au long de ces quelques mois,

Catherine, Christiane, Caroline, Danielle, Anélia et Yves, merci à vous pour votre aide et vos remarques constructives,

Tous les membres du laboratoire, merci à vous pour votre bonne humeur quotidienne,

Enfin, il m'est impossible d'oublier mes plus fervents supporters, mes parents. Merci à vous pour votre soutien et vos nombreux encouragements.

Table des matières

I.Introduction	1
1. Aperçu du développement du cerveau chez les vertébrés	1
2. Mutations affectant le développement cortical	4
3. Mutation <i>reeler</i> et le gène <i>reelin</i>	6
4. La voie de signalisation <i>reelin</i>	9
La régulation de l'expression	11
Îlots CpG	12
Caractérisation des îlots CpG	13
5. Régulation de l'expression d'un gène par la méthylation	14
6. But du travail	15
 II. Matériel et méthodes	 16
1. Souche bactérienne et plasmides	16
1.1 Souche bactérienne	16
1.2 Plasmides utilisés	16
1.2.1 pBluescript	16
1.2.2 pGL3-Basic	17
1.2.3 pCR2.1	17
1.2.4 pPAC4	17
2. Préparation d'ADN plasmidique	18
2.1 Clones provenant d'une banque génomique PAC	18
2.2 Clones provenant de phagémide	18
2.2.1 Midipréparation d'ADN	18
2.2.2 Minipréparation d'ADN	18
3. Quantification d'ADN	19
4. Réaction de polymérisation en chaîne (ou PCR)	20

4.1 Conditions classiques de PCR	20
4.2 Conditions particulières de PCR	21
PCR longue	21
5. Electrophorèse d'ADN sur gel d'agarose	22
6. Purification des produits PCR en vue du séquençage	23
7. Séquençage	23
7.1 Séquençage sur de l'ADN de phagémide	24
7.2 Séquençage sur de l'ADN de PAC	24
8. Criblage d'une banque d'ADN génomique	25
9. Electrophorèse d'ADN dans un champ électrique pulsé	26
10. Extraction d'ADN génomique de souris	26
11. Southern blot et marquage radioactif	27
12. Constructions génétiques et transformation de cellules	28
12.1 Constructions génétiques	28
12.1.1 Restriction de l'ADN	28
12.1.2 Déphosphorylation	28
12.2.3 Réaction de ligation	29
12.1.4 Réaction de « fill in » et de rognage	30
12.2 Transformation bactérienne	30
12.2.1 Transformation bactérienne avec les constructions génétiques	30
12.2.2 Transformation bactérienne avec un produit PCR cloné à l'aide du kit TA cloning (Invitrogen)	32
13. Mise en collection de bactéries	33
14. Transfection	33
14.1 Les plasmides	33
14.1.1 Le plasmide test	33
14.1.2 Le plasmide contrôle	33

14.2	Passage des cellules	33
14.3	Transfection par la Lipofectamine	34
14.4	Préparation du lysat cellulaire	35
14.5	Méthode de détection : le système dual-Luciférase	35

15.	Analyse informatique	36
-----	----------------------	----

III. Résultats 37

1.	Caractérisation de la région promotrice du clone λ 3-7	37
2.	Analyse des séquences et comparaison souris-homme	38
	Ilots CpG	38
	Sites potentiels de fixation de facteurs transcriptionnels	38
3.	Cribage d'une banque d'ADN génomique de souris	39
	Caractérisation du clone positif	40
	a. Taille de l'insert	40
	b. Contenu du PAC	40
4.	Etude de la régulation de l'expression par la méthylation	41
5.	Etude <i>in vitro</i> du promoteur <i>reelin</i> souris	42

IV. Conclusions, discussion et perspectives 47

V. Bibliographie 52

Annexes

- Annexe 1 Liste des amorces spécifiques des vecteurs utilisés
- Annexe 2 Liste des amorces spécifiques du gène *reelin*
- Annexe 3 Séquence du promoteur *reelin* souris avec traduction
- Annexe 4 séquence du promoteur *reelin* humain avec traduction

Introduction

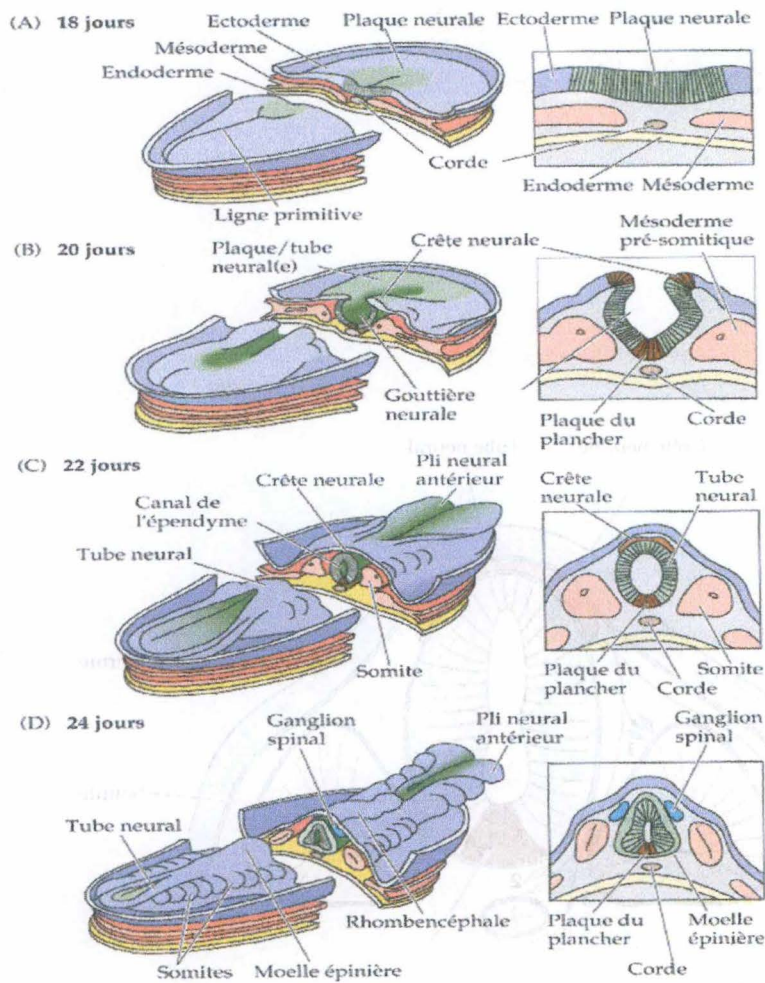


Fig.I.1: Neurulation de l'embryon de mammifère (Purves et al., 1999).

A gauche, vues dorsales de l'embryon à différents stades du développement précoce. A droite, coupes transversales passant par le milieu d'un embryon de même âge. (A) A la fin de la gastrulation et au début de la neurulation, la corde se forme par invagination du mésoderme dans la région de la ligne primitive. (B) A mesure que la neurulation progresse, la plaque neurale commence à se replier, formant la gouttière neurale. La partie de la plaque neurale située juste au-dessus de la corde se différencie en plaque du plancher. Sur les bords de la plaque neurale s'individualisent les crêtes neurales. (C) Les bords de la plaque neurale se soudent sur la ligne médiane pour former le tube neural. De chaque côté du tube neural, le mésoderme s'épaissit et se divise en structures, appelées somites, qui sont les précurseurs de la musculature axiale et du squelette. (D) Ensuite, le tube neural situé près des somites forme l'ébauche de la moelle épinière, les crêtes neurales donnant naissance aux ganglions spinaux. Enfin, les extrémités antérieures du tube neural (plis neuraux antérieurs) se rejoignent sur la ligne médiane et continuent à s'étendre pour donner naissance au cerveau.

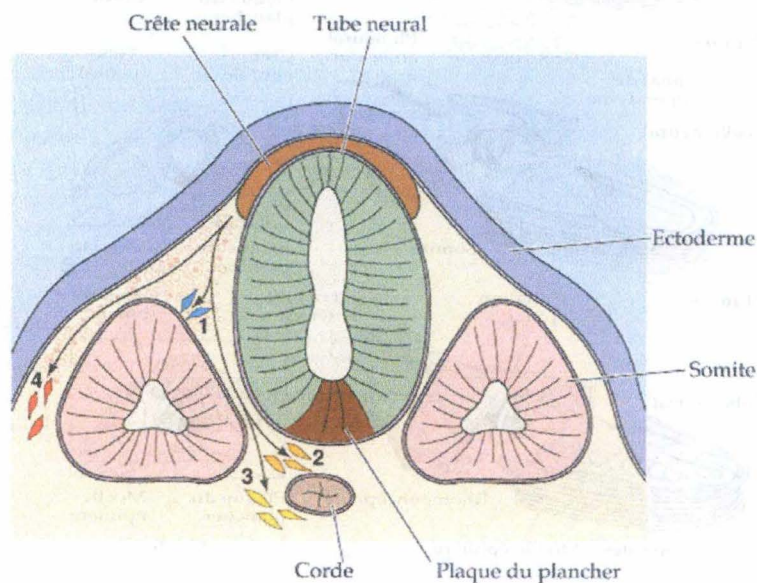


Fig.I.2: Schéma d'une coupe transversale d'un embryon de mammifère à un stade de développement correspondant à celui de la figure I.1.c (Purves et al., 1999). Les crêtes neurales migrent en suivant quatre voies différentes conduisant à la différenciation de structures et de types cellulaires distincts. Chaque voie offre des possibilités d'interaction avec différents environnements cellulaires dont les cellules migrantes reçoivent des signaux inducteurs (Sanes, 1989).

I. Introduction

1. Aperçu du développement du cerveau chez les vertébrés

Les trois feuillets de l'embryon se forment lors de la gastrulation: ectoderme, mésoderme et endoderme. Une conséquence de la gastrulation est la formation de la corde dorsale (ou notochorde), cylindre de cellules mésodermiques s'étendant de l'avant à l'arrière de l'embryon, sur la ligne médiane. La corde se forme à partir d'une masse mésodermique qui s'invagine et s'étend vers l'avant à partir d'une dépression superficielle, la fossette primitive, qui s'allonge pour former la ligne primitive. L'ectoderme situé juste au-dessus de la corde se transforme en plaque neurale, suite au processus appelé « induction neurale ». Ensuite, les bords de la plaque se replient vers l'intérieur, la transformant progressivement en gouttière puis en tube neural. De ce tube neural dérive par la suite la totalité du cerveau et de la moelle épinière. (Fig.I.1)

Les cellules précurseurs neurales sont des cellules souches qui se divisent et produisent d'autres précurseurs en nombre accru ainsi que des cellules qui ne se divisent plus, mais se différencient en neurones. Du fait de leur proximité par rapport à la corde, les cellules situées sur la ligne médiane ventrale du tube neural se différencient en une plaque basale. Les signaux inducteurs en provenance de la plaque basale entraînent une différenciation des cellules de la partie ventrale du tube neural aboutissant à la formation de motoneurones. Les cellules précurseurs qui sont plus éloignées de la ligne médiane donnent naissance aux neurones sensitifs des cornes postérieures. À la limite dorsale du tube neural, une troisième population de cellules fait saillie puis se détache pour former la crête neurale. (Fig.I.2). Les cellules de la crête neurale migrent à partir du tube neural en suivant quatre voies conduisant à la différenciation de structures et de types cellulaires distincts. Les cellules qui empruntent les voies numérotées 1 et 2 donnent respectivement naissance aux ganglions spinaux et sympathiques. Les précurseurs des cellules neurosécrétrices de la médullo-surrénale empruntent la voie migratoire n°3, tandis que celles qui sont destinées à former des tissus non nerveux empruntent la voie n°4. Chaque voie offre des

possibilités d'interaction avec différents environnements cellulaires dont les cellules migrantes reçoivent des signaux inducteurs.

Peu après la formation du tube neural, les ébauches des principales régions du cerveau prennent forme sous l'effet de mouvements morphogénétiques qui courbent, plissent et resserrent le tube neural. L'extrémité antérieure du tube se courbe en une sorte de crosse qui jouxte la courbure la plus prononcée, la flexion céphalique, se renfle et donne le cerveau antérieur ou prosencéphale. Le mésencéphale ou cerveau moyen naît de la dilatation de la flexion céphalique. Quant au cerveau postérieur ou rhombencéphale, il se forme à partir de la portion longue et relativement droite qui va de la flexion céphalique à la flexion cervicale. En position plus caudale, le tube neural forme l'ébauche de la moelle épinière. Sous l'effet de ces courbures et de ces plissements, la lumière du tube neural se modifie elle aussi et forme, au terme du développement, les ventricules cérébraux.

Après la mise en place des régions cérébrales primitives, celles-ci subissent encore au moins deux séries de segmentations pour former les régions du cerveau adulte. C'est ainsi que le prosencéphale rostral forme le télencéphale, qui comprend les ébauches du cortex cérébral, de l'hippocampe, des ganglions de la base, des noyaux du télencéphale basal et des bulbes olfactifs. La partie caudale du prosencéphale forme le diencéphale, d'où dérivent le thalamus et l'hypothalamus ainsi qu'une paire de boursofflures latérales (les cupules optiques) à partir desquelles se forme la partie nerveuse de la rétine. Enfin, la partie rostrale du rhombencéphale donne le métencéphale (d'où sont issus le cervelet et le pont du cerveau adulte), tandis que sa partie caudale donne le myélencéphale (futur bulbe).

À maturité, le cerveau humain comporte environ 1000 milliards de neurones et un nombre plus grand encore de cellules gliales. Les neurones se forment en quelques mois, à partir d'une population de cellules précurseurs situées dans la zone ventriculaire, une couche cellulaire profonde entourant la lumière du tube neural. Cette zone ventriculaire connaît une intense activité mitotique. Les cellules précurseurs passent au cours de leur cycle mitotique par une séquence stéréotypée de mouvements cellulaires et donnent naissance soit à

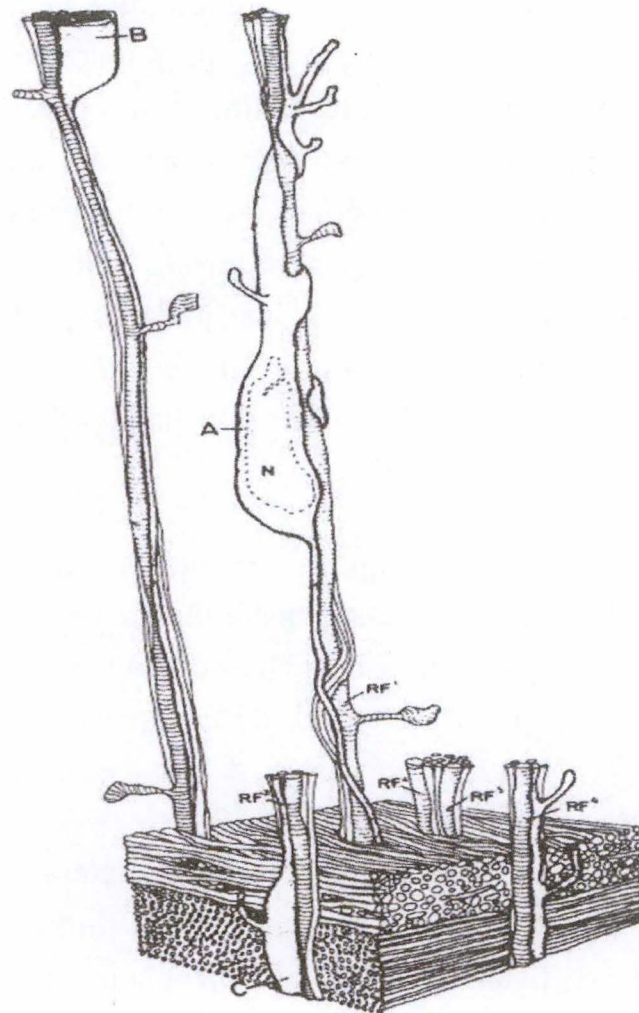


Fig.I.3: Représentation d'un neurone en migration le long d'une fibre gliale radiaire dans la zone intermédiaire (D'après Rakic, 1972).

de nouvelles cellules souches, soit à des cellules postmitotiques qui se différencient en neurones. Les neurones postmitotiques quittent la zone ventriculaire et migrent jusqu'à leur emplacement définitif dans le cerveau en développement.

Le fait de savoir à quel moment « naissent » les neurones destinés à une région cérébrale déterminée (c'est-à-dire à quel moment ils deviennent postmitotiques) a apporté des éclaircissements considérables sur la façon dont se constituent les régions du cerveau. Dans le cortex cérébral, par exemple, on a pu montrer que les six couches du cortex se forment de l'intérieur vers l'extérieur. Les premières cellules sont situées dans les couches les plus profondes, tandis que les neurones des générations ultérieures migrent au travers des cellules plus anciennes pour se disposer plus en surface.

Il existe plusieurs variétés de migrations neuronales. Les cellules des crêtes neurales migrent à travers le mésoderme et sont guidées par des molécules d'adhérence spécialisées de la matrice extracellulaire. À l'intérieur du système nerveux central, par exemple dans le cortex, le cervelet, l'hippocampe et la moelle, une bonne partie des neurones se dirigent vers leur destination finale en rampant le long d'un type particulier de cellule gliale, la glie radiaire, qui sert de guide cellulaire. Il existe aussi des cas de migration tangentielle où les cellules se déplacent presque perpendiculairement aux fibres gliales radiaires. Enfin, il existe une migration dite « en chaîne » qui conduit des neurones engendrés le long des ventricules vers les bulbes olfactifs. Les rapports entre migration en chaîne et tangentielle restent mal définis.

Au niveau du cortex cérébral, les premiers neurones postmitotiques (douzième jour embryonnaire, E12 chez la souris) quittent la zone ventriculaire et forment un réseau de neurones horizontaux appelé « préplaque ». Cette préplaque inclut des neurones de la zone marginale (future couche corticale ou moléculaire) comme les cellules de Cajal-Retzius, et les futures cellules de la sous-plaque. Bien que la majorité des neurones de la préplaque migrent radialement à partir de la zone ventriculaire (Fig.I.3), il est possible qu'une partie des cellules de Cajal-Retzius suivent une route tangentielle.

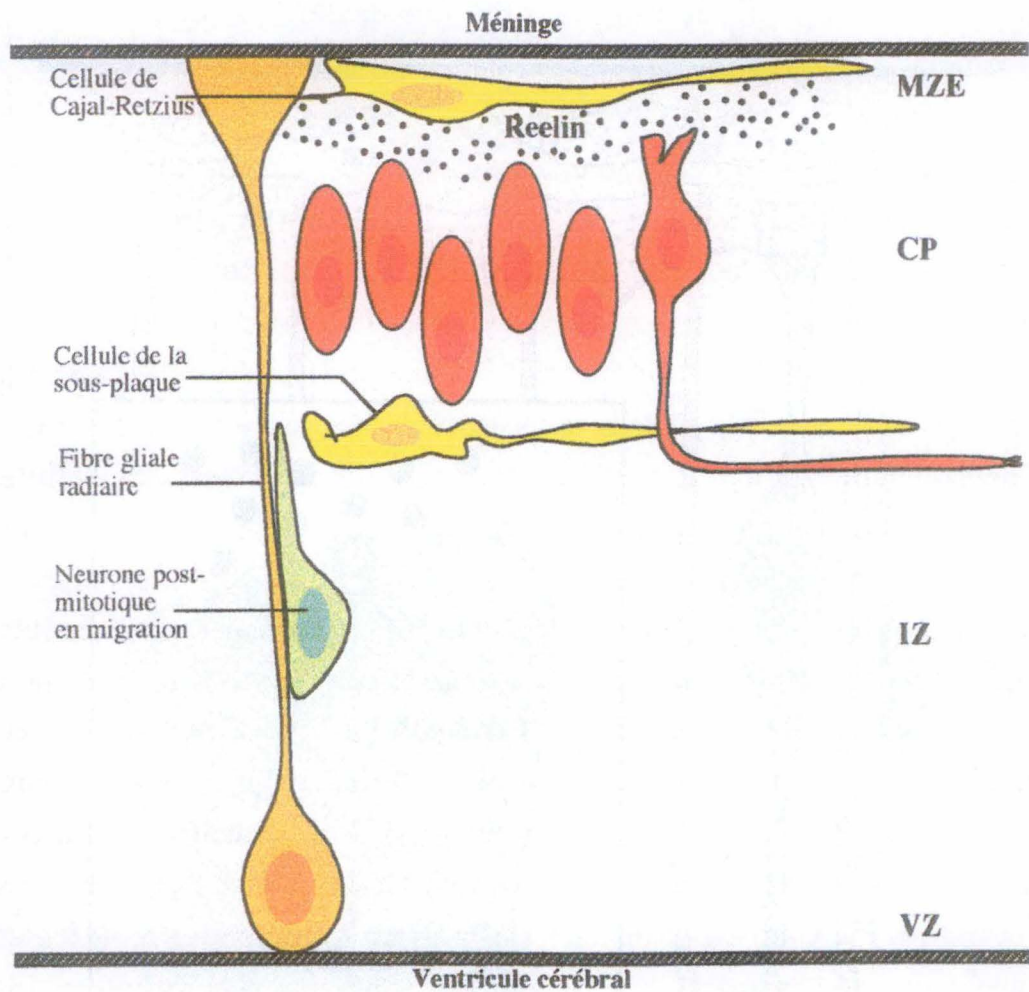


Fig.I.4: Schéma d'une coupe au niveau du cortex de souris au quatorzième jour embryonnaire. Le ventricule est en bas et les méninges en haut. Vers le quatorzième jour embryonnaire, les neurones postmitotiques nés dans la zone ventriculaire (VZ) migrent radialement le long des fibres gliales radiaires à travers la zone intermédiaire (IZ) vers la plaque corticale (CP). L'organisation horizontale primitive (en jaune) comprend les neurones de Cajal-Retzius, situés dans la zone marginale externe (MZE), et les cellules de la sous-plaque. Ces deux contingents sont séparés par la plaque corticale dont les neurones ont une organisation radiales (en rouge). La *reelin* est sécrétée par les neurones de Cajal-Retzius.

L'étape suivante est la formation de la plaque corticale (E14 chez la souris) constituée en majorité de neurones migrant le long de fibres radiales. Les premières cellules qui migrent ont une forme étoilée tandis que les plus jeunes ont une forme allongée bipolaire et sont plus « fermement » reliées aux fibres gliales radiales. De plus, certains neurones générés dans l'éminence ganglionnaire, atteignent la plaque corticale par migration tangentielle.

La condensation de la plaque corticale divise la population cellulaire de la préplaque en deux contingents. Une partie des cellules de la préplaque se situent dans la zone marginale externe, tandis que l'autre partie se trouvent déplacées en dedans de la plaque et forment la « sous-plaque », la future couche corticale VIb. La plaque corticale est le précurseur de la plupart du cortex et se développe par addition d'éléments migrant le long de fibres radiales à partir de la zone ventriculaire avant de s'organiser en couches de plus en plus superficielles (Fig.I.4). Les couches corticales profondes, V et VIa, sont engendrées les premières, alors que les neurones plus jeunes forment progressivement les couches IV, III, et II. Ce gradient corticogénétique « inside-out » est une caractéristique générale du cortex de tous les mammifères, y compris des marsupiaux. Ce schéma classique de migration radiale vers le cortex s'applique aux neurones excitateurs glutamatergiques qui forment environ 80 % de la population corticale. Ces dernières années, il a été montré que les neurones GABAergiques inhibiteurs, soit environ 20 %, proviennent de zones ventriculaires localisées en regard du striatum, dans l'éminence ganglionnaire, et que ces neurones gagnent le cortex par migration tangentielle.

2. Mutations affectant le développement cortical

Durant de nombreuses années, la mutation *reeler* chez la souris (mutation autosomique récessive du gène *reelin*) était la seule mutation connue, à toucher le développement cortical, mais d'autres ont été décrites plus récemment. Le même phénotype que *reeler* est engendré par des mutations du gène *Disabled1* (*Dab1*), par exemple chez les souris KO pour *Dab1* ou dans deux mutations spontanées de *Dab1* appelées *scrambler* et *yotari*. De même, le phénotype *reeler* est également présent chez des souris qui sont doublement mutantes (double KO) pour deux gènes qui codent des récepteurs de lipoprotéines, à savoir Very

Low Density Lipoprotein Receptor (VLDLR) et Apolipoprotein E Receptor type 2 (ApoER2) (Hiesberger et al., 1999). Ces mutations définissent une nouvelle voie de signalisation impliquée dans le développement cortical (cette voie est détaillée au point 4).

Un autre mutant de souris dont le phénotype ressemble à *reeler*, affecte le gène codant pour une sérine/thréonine kinase exprimée dans les neurones postmitotiques, la «cyclin dependent protein kinase 5 » (Cdk5 ; Tsai et al., 1993). Cette mutation touche la migration radiale, rendant impossible la formation corticale « inside-out ». Une déficience du gène codant pour un cofacteur activateur (p35) de la kinase Cdk5 engendre aussi une anomalie du gradient histogénétique du cortex (Chae et al., 1997).

Toujours chez la souris, une déficience du facteur transcriptionnel Tbr1 donne un cortex au développement inversé ainsi qu'un déficit en cellules de Cajal-Retzius et en *reelin* (Hevner et al., 1999), alors que l'inactivation du gène p73 conduit aussi à une diminution de la synthèse de *reelin* et à une anomalie de l'hippocampe (Zhang et al., 2000). Enfin, des mutations des gènes Pax6 et Emx2 affectent aussi à des degrés divers le développement du cortex (Bishop et al., 2000 ; Mallamaci et al., 2000).

Chez l'homme, certaines maladies génétiques perturbent la migration des neurones corticaux. Il s'agit surtout de la lissencéphalie de type 1 ou vraie, de la malformation dite « double cortex », et des hétérotopies périventriculaires liées au chromosome X. Dans la lissencéphalie de type 1 (forme Miller-Dieker), le cortex cérébral est dépourvu de circonvolutions (Reiner et al., 1995). Le gène muté, appelé *LIS1*, code pour une protéine qui interagirait avec les microtubules. La malformation appelée « double cortex » est caractérisée par deux populations de neurones, une normale et l'autre dont la migration est défective. On la décrit chez des filles, et des garçons affectés de lissencephalie naissant dans les mêmes familles. Le gène responsable est situé sur le chromosome X et la malformation « double cortex » reflète l'inactivation du X chez les filles porteuses. Le gène est appelé « *doublecortin* » et semble aussi interagir avec les microtubules (revue dans Walsh et Goffinet, 2000). Les hétérotopies périventriculaires ou nodulaires se caractérisent par une population de neurones qui n'ont pas achevé leur migration radiale normale. Dans un type hérité de manière liée au sexe, la

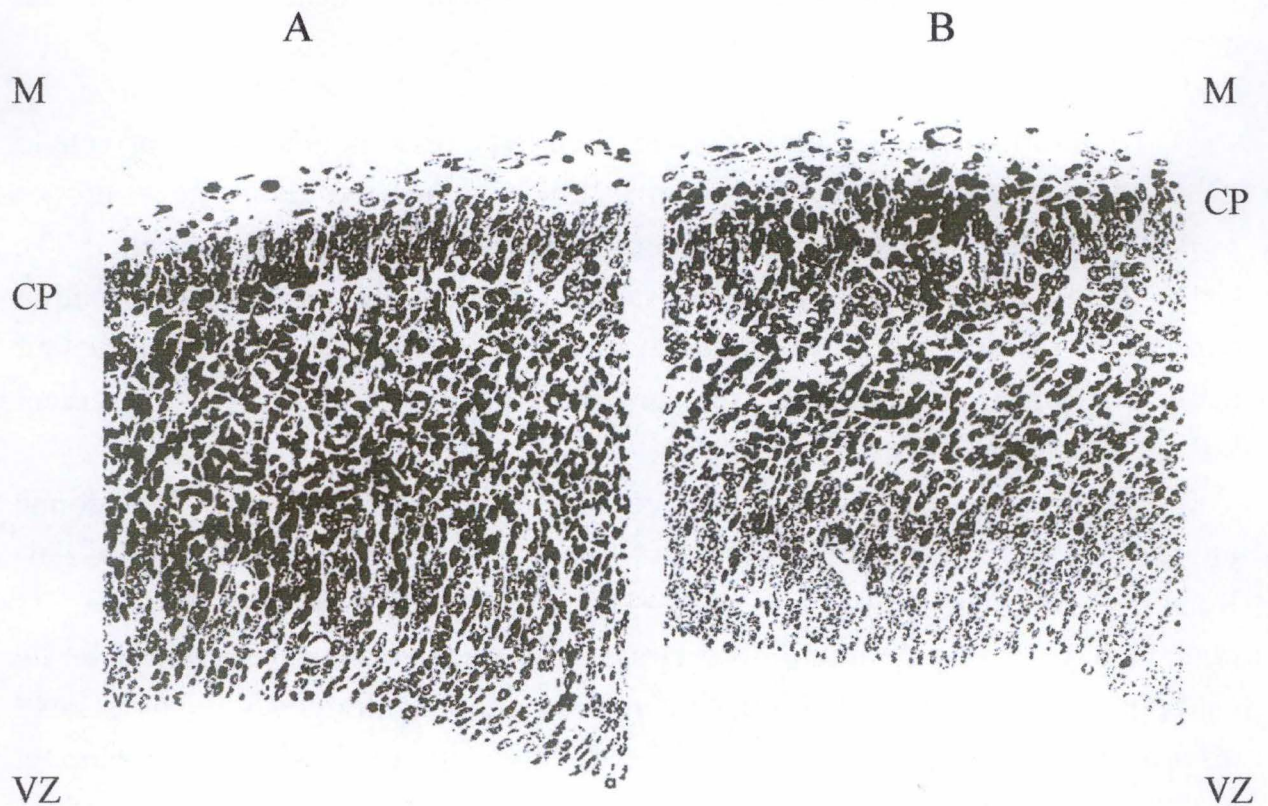


Fig.I.5: Coupes semi-fines pratiquées au niveau du cortex embryonnaire normal (A) et *reeler* (B) au quatorzième jour.

Les neurones sont engendrés dans les zones ventriculaires (VZ) situées le long des ventricules et migrent à travers la zone intermédiaire en suivant les fibres gliales radiaires, en direction des méninges (M). Chez l'animal normal (à gauche), la plaque corticale (CP) possède une organisation radiaire et définit deux zones pauvres en cellules appelées zone marginale et sous- plaque corticale. Chez l'embryon *reeler* (à droite), la plaque corticale est moins bien organisée.

mutation touche le gène de la filamine, une protéine associée à l'actine. On voit que les mutations de la migration caractérisées chez l'homme, convergent toutes vers le cytosquelette (Lambert de Rouvroit et Goffinet, 1998).

Récemment, une pathologie associée à une mutation du gène *reelin* humain a été découverte par le laboratoire de C. Walsh. Il s'agit d'une forme récessive autosomique de lissencéphalie associée à une sévère hypoplasie du cervelet. La mutation du gène *reelin* en modifie l'épissage et ressemble à l'allèle *reeler albany2* (Royaux et al., 1997 ; Lambert de Rouvroit & Goffinet, 1998).

3. La mutation *reeler* et le gène *reelin*

Reeler est une mutation récessive autosomique isolée par Falconer en 1951 et cartographiée sur le chromosome 5 ; depuis le clonage du gène affecté, baptisé *reelin*, le symbole est *Reln* (d'Arcangelo et al., 1995).

Les mutants *reeler* présentent des anomalies cytoarchitectoniques dans plusieurs régions cérébrales, surtout dans le néocortex, l'hippocampe et le cervelet (Fig.I.5). Chez l'embryon *reeler*, les neurones sont engendrés en nombre et selon une chronologie normale, et migrent le long des fibres gliales radiaires comme chez les embryons normaux. Cependant, arrivés près de leur destination, les neurones en migration ne parviennent pas à s'ordonner selon les agencements architectoniques normaux. Le phénotype est le plus évident au niveau du cervelet, qui est atrophié et dont le cortex est pratiquement lisse. L'anomalie cérébelleuse est à l'origine de l'ataxie qui permet l'identification des animaux atteints. L'analyse du développement cérébelleux montre que le déficit primaire est une incapacité des cellules de Purkinje de former une plaque bien organisée. Le néocortex et l'hippocampe sont aussi très anormaux avec une mauvaise définition des couches neuronales et l'absence d'une couche moléculaire (couche I) bien définie. Les neurones de la plaque corticale sont mal orientés et déplacent vers la zone marginale l'ensemble des éléments de la préplaque. Tant les cellules de Cajal-Retzius que celles de la sous-plaque se retrouvent à la partie externe, dans la zone marginale. De plus, les neurones en migration vers la plaque corticale ne traversent pas les couches préalablement établies mais se déposent à des niveaux progressivement plus profonds, de sorte que le gradient de maturation du cortex est inversé, dirigé de dehors en dedans

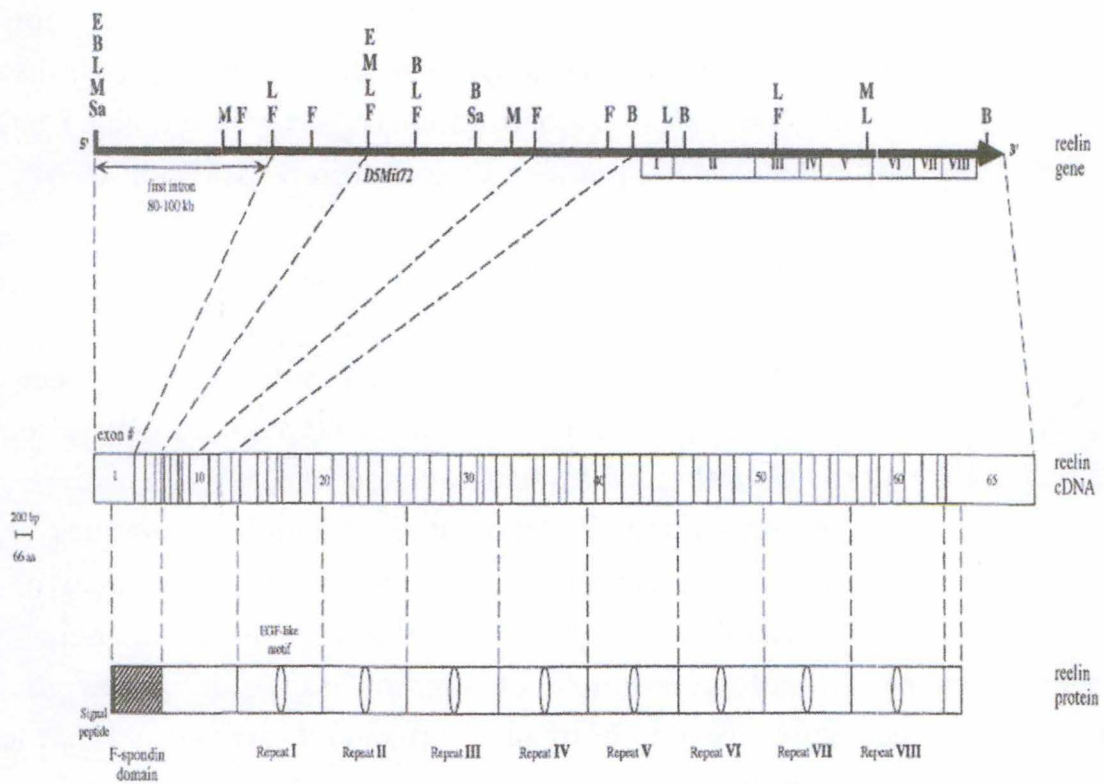


Fig.I.6: A: Le gène *reelin* contient le microsatellite *D5Mit72*. Les enzymes de restriction suivants: E= *EagI*; B= *BssHII*; L= *SalI*; M= *MluI*; Sa= *SacII*; F= *SfiI*, coupent fréquemment dans les îlots CpG.
 B: Structure de l'ADNc de *reelin*; les exons sont représentés par des boîtes.
 C: Organisation de la protéine *reelin*.

« outside-in ». De nombreuses autres régions du système nerveux sont affectées, comme l'olive bulbaire ou les noyaux cochléaires, de sorte que l'ensemble des neurones centraux sont touchés à des degrés divers par la mutation. À tous ces niveaux, la lésion élémentaire est semblable et consiste en un déficit de formation des profils architectoniques en fin de migration. Le système nerveux périphérique ainsi que les autres organes ne sont apparemment pas affectés (Bernier et al., 1998).

Le gène *reelin*, dont les mutations engendrent le phénotype *reeler*, est un gène de grande taille, entre 400 et 450 kpb. Il contient 65 exons et code pour un messager de quelque 12 kb (Fig.I.6). Les exons de la partie 5' du gène sont généralement plus petits que ceux des parties centrale et terminale. Les introns de la partie 5' sont en revanche très grands, en particulier le premier qui est énorme, de l'ordre de 80 kbp ; la taille de cet intron est de 72 kbp chez l'homme. Par conséquent, la partie 5' du message est étalée sur une grande région d'ADN, alors que les répétitions sont codées par un bloc de moins de 200 kpb. Les grands introns pourraient avoir comme fonction d'allonger le temps de transcription d'un gène impliqué à des stades précis du développement. Ils pourraient aussi contenir des séquences importantes pour la régulation de l'expression du gène. Toutes les jonctions exon-intron se conforment à la règle GT/AG, à l'exception d'une jonction GC/AG ; aucun intron du type AT/AC n'a été mis en évidence (Royaux et al., 1997).

Une recherche systématique de formes alternatives de la *reelin* a mis en évidence deux événements impliquant l'extrémité 3' du gène. Le premier concerne l'exon 64, absent dans 10 % environ des transcrits *reelin* du cerveau. Cet exon ne comporte que 6 nucléotides et code pour deux acides aminés Val-Ser. Le second consiste en l'utilisation d'un site alternatif de polyadénylation ayant pour effet d'introduire des codons stop dans le cadre de lecture amputant la protéine prédite d'une séquence fortement basique de 33 acides aminés (Royaux et al., 1997 ; Lambert de Rouvroit et al., 1999a). La forme 3' alternative représente 10 à 25 % de l'ensemble des ARNm *reelin* à tous les stades du développement (Bernier et al., 1998). Bien que le rôle physiologique de ces événements d'épissage alternatif reste inconnu, il faut souligner qu'ils sont

présents chez la souris, le rat, le poulet, l'homme, et que semblable conservation évolutive suggère une fonction importante.

Reelin est une protéine de 3.461 acides aminés dont la taille prédite est de 388 kDa. Il s'agit d'une glycoprotéine sécrétée dans la matrice extracellulaire (d'Arcangelo et al., 1997). La partie N-terminale débute par un segment hydrophobe de type peptide signal, suivi par un segment de 190 résidus ressemblant à la F-spondine (25 % d'identité), une protéine sécrétée par la plaque basale du tube neural embryonnaire. Un segment unique de 500 résidus est suivi de 8 répétitions de 350 à 390 résidus. Chacune de ces répétitions comprend un motif EGF (epithelial growth factor) semblable à ceux de la ténascine, flanqué de deux segments A et B qui ont également une certaine similitude entre eux. La protéine se termine par une séquence fortement basique de 33 résidus (Fig.I.6). Dans le cerveau, la *reelin* n'est pas intacte mais clivée en deux sites par une activité protéase de type métalloprotéinase (Lambert de Rouvroit et al. 1999b). Ce processing de la protéine est aussi conservé dans plusieurs espèces et revêt probablement une importance fonctionnelle qui reste à définir.

Reelin est détectable dans le cerveau embryonnaire de souris à partir du 12ème jour (E12). Sa concentration augmente au cours du développement, culmine autour de la naissance et décroît ensuite graduellement pour atteindre un faible niveau résiduel chez l'adulte. À tous les stades embryonnaires et jusqu'à quelques jours après la naissance, la synthèse de la *reelin* est extrêmement forte dans les cellules de Cajal-Retzius, ainsi qu'au niveau des neurones correspondants de la zone marginale dans l'hippocampe, dans les cellules mitrales du bulbe olfactif depuis les stades précoces jusqu'à l'âge adulte, dans la rétine au niveau des cellules ganglionnaires et enfin dans les grains cérébelleux au stade précoce. L'ARNm de *reelin* est indétectable dans plusieurs structures clairement anormales chez le mutant, comme la plaque corticale, les cellules de Purkinje, l'olive bulbaire ou encore le noyau du nerf facial (Schiffmann et al., 1997).

Les observations sur l'expression de l'ARNm et de la protéine *reelin* montrent qu'il n'existe pas de corrélation entre l'expression de *reelin* et le

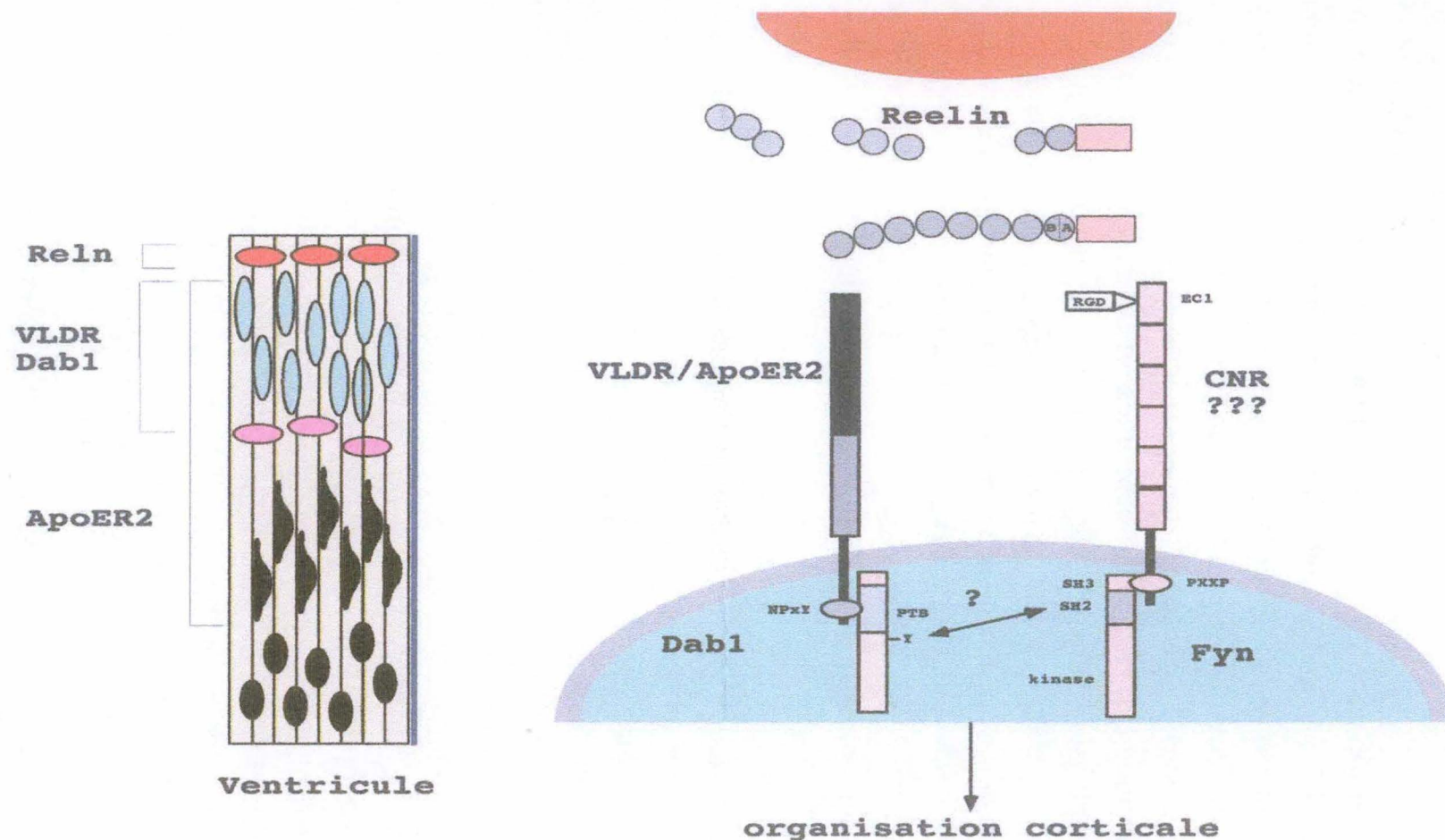


Fig.I.7: La voie de signalisation *reelin*.

A: Les neurones Cajal-Retzius (en rouge) expriment *reelin*. Les neurones de la plaque corticale (en vert) expriment les gènes *VLDR* et *Dab1*. Les neurones de la sous-plaque sont en rose. Les neurones post-mitotiques en migration (forme allongée de couleur noire), ainsi que les neurones de la plaque corticale et de la sous-plaque, expriment *ApoER2*.

B: Les neurones de Cajal-Retzius (en rouge) expriment *reelin*. Cette dernière interagit avec un neurone de la plaque corticale (en vert), exprimant les récepteurs *VLDR*, *ApoER2* et *CNR*, ainsi que l'adaptateur *Dab1* et la kinase *Fyn*.

phénotype *reeler*. Cette discordance s'explique par une action « paracrine » de la *reelin*, qui est sécrétée par certains neurones mais agit à distance sur des neurones cibles qui sont différents des neurones producteurs. En présence de *reelin*, les neurones en migration perçoivent un signal qui les pousse à se détacher des fibres gliales radiaires et à prendre une disposition architectonique spécifique (Schiffmann et al., 1997). Le mécanisme par lequel le signal *reelin* est traduit par les neurones cibles commence à être connu, bien que de nombreux points imprécis subsistent.

4. La voie de signalisation *reelin*

Comme mentionné plus haut, un phénotype identique à *reeler* est engendré aussi par des mutations du gène *Dab1*, ainsi que par deux mutations simultanées des récepteurs de lipoprotéines « very low density lipoprotein receptor » (VLDLR) et « apolipoprotein E receptor-2 » (ApoER2) (Trommsdorf et al., 1999). Tant *Dab1* que VLDLR et ApoER2 sont exprimés par les neurones cibles, c'est-à-dire ceux qui sont touchés par la mutation *reeler*. Ceci suggère que ces différentes protéines sont impliquées dans une même voie de régulation de la migration neuronale et définissent une nouvelle voie de signalisation dont le ligand est la *reelin*, les récepteurs VLDLR et ApoER2, et *Dab1* un adaptateur intracellulaire nécessaire. La liaison directe de *reelin* sur les récepteurs VLDLR et ApoER2 a été démontrée (Hiesberger et al., 1999 ; d'Arcangelo et al., 1999). Ces récepteurs VLDLR et ApoER2 possèdent une séquence cytoplasmique NPxY qui se fixe au domaine PI/PTB présent à la partie N-terminale de l'adaptateur *Dab1* (Fig.I.7). D'autre part, plusieurs résidus de *Dab1* peuvent, lorsqu'ils sont phosphorylés sur certains résidus Y, être reconnus par le domaine SH2 des tyrosines kinases de la famille Src (par exemple ; Src, Fyn, Yes, Abl). Lorsque les résidus Y de la protéine *Dab1* sont mutés in vivo (technique dite de « knock-in »), le phénotype mutant est complet, de sorte que ces résidus Y sont nécessaires à l'activité biologique de *Dab1* (J. Cooper, communication personnelle). L'ajout de *reelin* à des neurones corticaux embryonnaires stimule la phosphorylation des résidus Y de *Dab1* (Howell et al., 1999), confirmant l'existence d'une cascade de signalisation biochimique.

Outre le fait que VLDLR et ApoER2 sont des récepteurs de *reelin*, une étude récente (Senzaki et al., 1999) montre que la protéine *reelin* est susceptible

d'interagir avec d'autres récepteurs, en particulier avec des protocadhérines de la famille CNR (cadherin-related neuronal receptors). Ces dernières possèdent six ou sept domaines extracellulaires ressemblant à ceux des cadhérines classiques. Cependant, leur partie cytoplasmique est différente et contient plusieurs motifs PxxP capable de se lier au domaine SH3 de la tyrosine kinase Fyn. Une interaction entre CNR1 et *reelin* nécessiterait la partie distale de la première répétition de *reelin* ainsi que le segment compris entre les acides aminés Ser-68 et Gly-97 du premier domaine extracellulaire de CNR1. Ce segment comprend un triplet RGD (motif connu pour les interactions protéine-protéine) dont le résidu D serait nécessaire à la liaison. Ces nouvelles données suggèrent donc que des protéines de la famille CNR se comportent aussi comme des récepteurs de *reelin*.

Enfin, il est possible que *reelin* se fixe sur une troisième famille de récepteurs, à savoir l'intégrine α -3 β -1 (Anton et al., 1999; Anton & Walsh, communication personnelle).

Les récepteurs CNR et VLDLR ou ApoER2 relaient le signal *reelin* dans la cellule de manière à modifier le phénotype par des voies qui restent à élucider. La réponse des neurones de la plaque corticale à *reelin* pourrait impliquer le cytosquelette neuronal via des molécules telles que Cdk5/p35, *doublecortin*, *Lis1*, filamine et d'autres (Walsh et Goffinet, 2000), et/ou modifier des propriétés d'adhésion telles que le détachement des fibres radiales et l'adhésion homotypique entre ces neurones.

Beaucoup de points restent à éclaircir. En effet, outre le manque d'arguments génétiques, les interactions de *reelin* avec les autres membres de la famille CNR doivent être étudiées en détail. De plus, seule la partie N-terminale de *reelin* a été testée et il est possible que d'autres sites de liaison existent à d'autres endroits de la protéine, pour des protéines CNR ou pour d'autres molécules. Qui plus est, les régions de *reelin* responsables de la liaison aux récepteurs VLDLR et ApoER2 doivent encore être définies, même si l'on sait déjà que la partie N-terminale impliquée dans l'interaction avec les CNR ne se lie pas aux récepteurs de lipoprotéines.

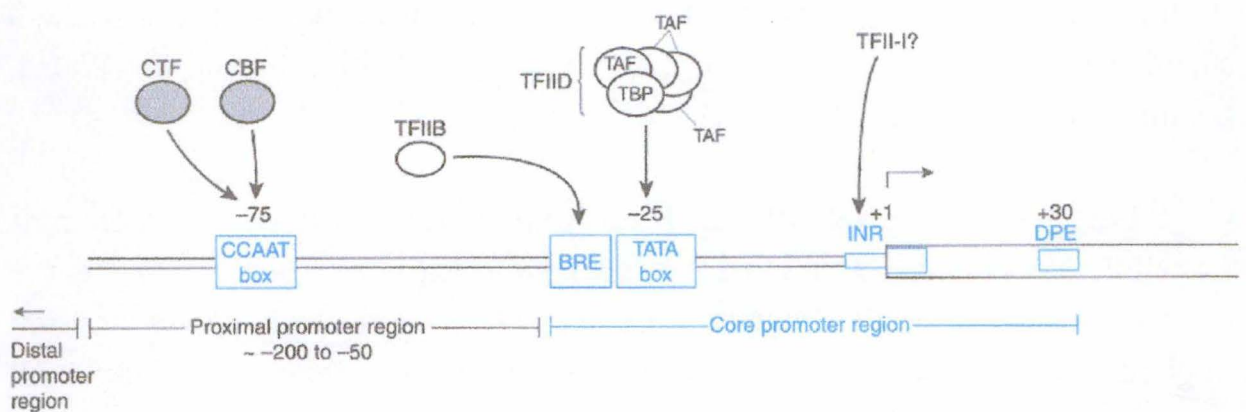


Fig.I.8: Localisation des sites conservés au niveau de la région promotrice des gènes eucaryotes. DPE; downstream promoter element, CTF, CCAAT-binding transcription factor, CBF; CCAAT box-binding factor, TBP; TATA box-binding protein, TAF; TBP-associated factors. Certains gènes n'ont pas tous les éléments. Par exemple, beaucoup de gènes ne possèdent pas de boîte TATA, Les gènes n'ont pas besoin de tous les éléments présents au niveau du cœur du promoteur. Par exemple, Les gènes n'ont pas besoin de tous les éléments présents au niveau du cœur du promoteur. Par exemple, les gène *reelin* ou $\beta 4$ -*Integrin* ne possèdent pas de boîte TATA, et le gène *reelin* ne possède aucun de ces éléments consensus.

La régulation de l'expression.

Dans le cas spécifique du cortex cérébral, la protéine *reelin* est produite par des cellules de la zone marginale, en particulier les neurones de Cajal-Retzius, alors que les récepteurs VLDLR et ApoER2 ainsi que l'adaptateur *Dab1* sont exprimés par les cellules de la plaque corticale. Cette expression différentielle pose la question de la régulation de l'expression des divers gènes de la cascade.

L'expression génétique est contrôlée par les promoteurs. Ces derniers sont des combinaisons d'éléments de courtes séquences localisés immédiatement en amont du site d'initiation de la transcription. Le cœur des promoteurs des gènes de mammifères contient des éléments qui, en absence d'éléments régulateurs additionnels, permettent une expression constitutive du gène, mais à un faible niveau. Ils sont généralement localisés entre -45 et +40 pb par rapport au site d'initiation. Il s'agit d'éléments comme la boîte TATA, dont la séquence consensus est TATA(A/T)A(A/T). Localisée à -25 pb, elle est entourée par des séquences riches en GC et reconnue par le complexe TFIID. La séquence BRE est située en amont de la boîte TATA et est reconnue par le complexe TFIIB. La séquence Inr ayant pour consensus PyPyCAPuPu (Py=pyrimidine, Pu=purine) est localisée au niveau du site d'initiation de la transcription. En fin, à +30 pb de ce dernier, on retrouve la séquence DPE (Downstream Promoter Element). Des séquences formant le promoteur proximal se situent entre -50 et -200 pb. Il s'agit par exemple des boîtes GC (consensus ; GGGCGG) et CCAAT (consensus ; GGCCAATCT) (Fig.I.8). Les promoteurs contiennent des sites de reconnaissance pour des facteurs régulateurs de la transcription.

Les « enhancers » sont des activateurs servant à augmenter le niveau de la transcription. Leur activité est indépendante de leur orientation et de leur distance par rapport au gène qu'ils régulent (Blackwood et Kadonaga, 1998). Les « silencers » servent à réduire le niveau de la transcription. On en distingue deux classes. Les silencers classiques, dont l'activité ne dépend pas de leur position, ont un mécanisme de répression actif, au contraire des éléments de régulation négative (Ogbourne et Antalis, 1998). Les « insulateurs » ou éléments

frontières sont des régions de l'ADN, longue de 0.5 à 3kb, dont la fonction est de bloquer l'influence d'agents ayant un effet positif ou négatif sur la transcription. Enfin, « les éléments de réponses » ont pour fonction de moduler la transcription en réponse à un stimuli spécifique externe. Ils sont généralement localisés à 1kb du site d'initiation de la transcription.

Lors d'un travail antérieur, environ 700 nucléotides situés en amont du site d'initiation de la traduction du gène *reelin* souris ont été déterminés. Cette séquence est dépourvue de boîte TATA et contient une région très riche en CG qui possède toutes les caractéristiques d'un îlot CpG (Royaux et al., 1997). La séquence complète (162912nt) d'un clone BAC humain contenant, entre autres, la région promotrice du gène *reelin* est accessible dans GenBank (AC002067). La comparaison des deux promoteurs souris et humain montre qu'ils sont tous deux dépourvus d'une boîte TATA typique. Dans le cas du promoteur *reelin* humain, l'îlot CpG s'étend dans le premier intron.

Îlots CpG

Les îlots CpG sont également nommés îlots HTF (Hpa Tiny Fragment) car ils contiennent de nombreux sites clivables par l'enzyme de restriction Hpa II, qui est sensible à la méthylation, ou encore MTFs pour Methylation Free Islands (Bird, 1987).

Ces îlots sont des séquences riches en doublets CG, généralement hypométhylés. Cependant, ils sont méthylés dans divers cas. Lors de l'inactivation du chromosome X, les îlots méthylés se trouvent sur le chromosome inactif ; les îlots sont également méthylés lors de l'empreinte génétique, dans les cellules tumorales, au niveau des sites fragiles (Vincent et al., 1991) ainsi que dans les lignées cellulaires en culture depuis un temps suffisamment long. Dans ce dernier cas, les îlots associés aux gènes dont la transcription est superflue pour la croissance cellulaire sont méthylés (Bickmore et al., 1992). Les îlots CpG se distinguent des autres régions de l'ADN qui ne présentent que 20 à 25 % de la fréquence attendue en ces dinucléotides. Dans un sens, les îlots CpG ne sont pas riches en CpG, car ces dinucléotides surviennent à une fréquence attendue. C'est plutôt la région de l'ADN « inter-îlot » qui est

anormalement pauvre en CpG. Cela est une conséquence de la méthylation sur la cytosine. En effet, la cytosine une fois méthylée, est sujette à la désamination. Cette mutation produit des thymines qui ne sont pas reconnues par l'uracile-ADN-glycosylase et ne sont pas réparées. Avec le temps, les îlots CpG sont remplacés par des dinucléotides AT (Bird, 1987).

Les îlots CpG sont rencontrés dans la plupart des gènes de maintenance (70 %) transcrits par l'ARN Pol II. Mais il n'est pas rare de les trouver associés à des gènes plus spécifiques, comme les gènes *reelin* et *Thy-1* de la souris, ou encore le gène α -globine chez l'humain. La grande majorité des îlots CpG sont présents au niveau de la région 5' du gène, mais dans quelques gènes à expression spécifique, on les retrouve en aval du site d'initiation de la transcription. Par exemple, dans le gène humain APOE, un îlot CpG de 940bp est présent dans l'exon 4 (Jones et al., 1999).

Du point de vue de l'évolution, on suppose que ces îlots représentent les derniers « survivants » de la très grande fraction non méthylée (70 à 100 %) constituant le génome des invertébrés. La raison pour laquelle la méthylation s'est étendue si rapidement au cours de l'évolution n'est pas claire ; de même, nos connaissances du processus déclencheur de la méthylation *de novo* de l'ADN non méthylé restent incomplètes (Jones et al., 1999).

Caractérisation des îlots CpG

Les îlots CpG sont définis comme des régions longues d'au moins 200 pb avec un contenu en G+C d'au moins 50 % dont au moins 60 % de CpG.

La formule permettant d'attribuer un score à un îlot CpG présumé est la suivante :

$$\text{CpG O/E} = (\text{Nombre de CpG} / \text{Nombre de C} \times \text{nombre de G}) \times N$$

Avec N = nombre total de nucléotides dans la séquence. (Gardinner – Garden et Frommer, 1987). Pour être qualifié d'îlot CpG, la séquence étudiée doit contenir au moins 50 % de G+C et la valeur CpG O/E supérieure ou égale à 0,6. Les valeurs des îlots CpG oscillent généralement entre 0,6 et 1,1.

5. Régulation de l'expression d'un gène par la méthylation

La méthylation de l'ADN est un aspect de la biologie moléculaire des vertébrés qui diffère de celle des invertébrés. En effet, dans le génome de ces derniers, la fraction méthylée est inférieure à 30 %, voir totalement absente chez certaines espèces (comme *Drosophila*). Par contre, dans le cas des vertébrés, cette fraction est de l'ordre de 99 % (Bird, 1987).

Riggs, Holliday et Pugh, proposèrent il y a déjà 25 ans, que la méthylation de l'ADN faisait partie d'un système de contrôle, concernant l'expression des gènes dans les cellules de mammifères (Holliday et Pugh, 1975). Le terme «contrôle» implique la réversibilité, cependant, des travaux ultérieurs ont montré que la fonction la plus probable de la méthylation de l'ADN, entraîne la suppression permanente de l'expression du gène à travers les modifications des îlots CpG, situés au niveau de la région promotrice.

Ce silencing permanent ne peut être supprimé, normalement, que lors du passage à travers la lignée germinale ou durant le stade précoce du développement, car il n'y a aucun exemple connu d'îlots CpG méthylés devenus non méthylés à des stades ultérieurs (Jones, 1999).

Les connaissances actuelles sont plus en accord avec l'idée que la méthylation des îlots CpG agit dans un seul sens, et inactive un ou deux allèles d'un gène (empreinte génétique, inactivation du chromosome X,...), ou inactive de manière permanente le promoteur durant toute la vie de l'organisme (« silencing » transcriptionnel de séquences parasites) (Jones, 1999 ; Walsh & Bestor, 1999).

6. But du travail

Comme nous l'avons résumé, le gène *reelin* détient un rôle central dans le développement cérébral embryonnaire. Sa structure génomique est connue ainsi que de nombreuses informations concernant sa voie de signalisation.

En revanche, on dispose de très peu de données sur la régulation de l'expression de *reelin* et notre laboratoire a décidé de l'étudier, en commençant par caractériser sa région promotrice qui est le sujet de notre travail.

Pour cela, nous allongerons tout d'abord le fragment de promoteur cloné au cours d'un travail antérieur (Royaux et al., 1997). Ensuite, diverses constructions contenant des parties du promoteur seront analysées par transfection transitoire dans une lignée cellulaire adéquate. En parallèle, nous criblerons une banque d'ADN génomique, dans le but d'avancer le plus loin possible dans la région 5'. Une fois notre clone obtenu, nous le caractériserons et le séquencerons.

Nous étudierons également le problème de la régulation de l'expression du gène *reelin* par la méthylation. Pour ce faire, nous analyserons le pattern de méthylation de l'îlot CpG de *reelin* à partir d'ADN génomique extrait de différents organes de souris.

Matériel et méthodes

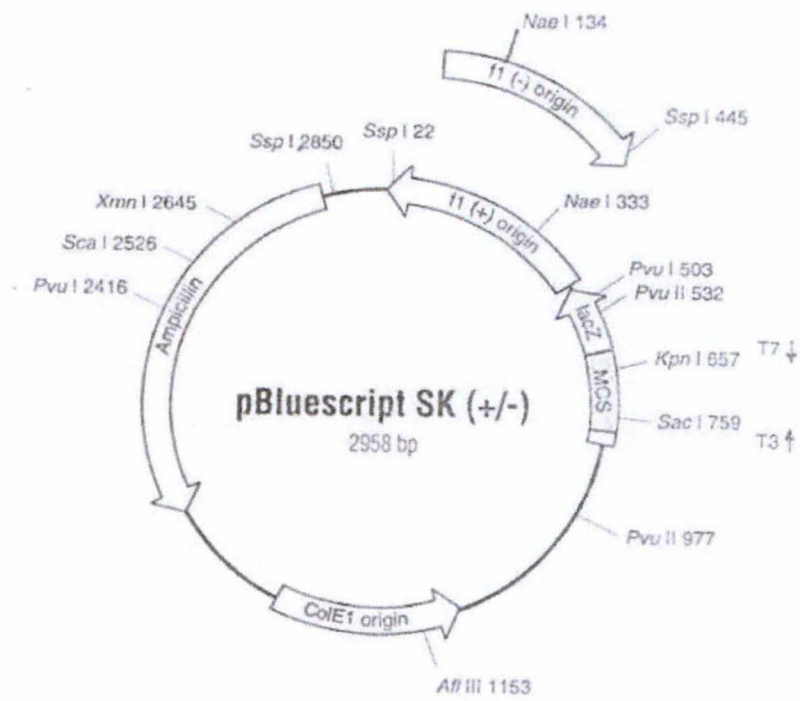


Fig.II.1: Vecteur pBluescript

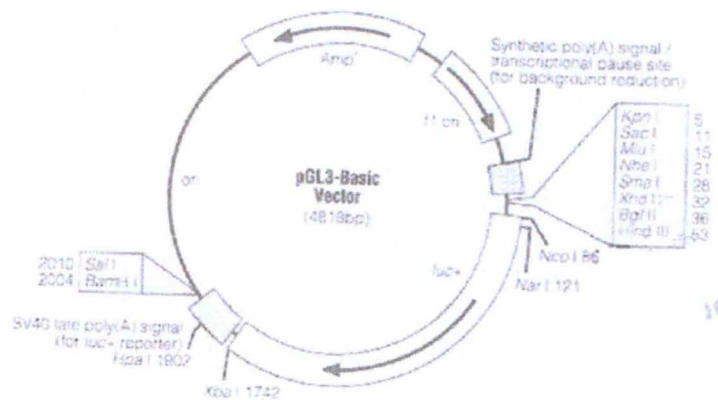


Fig.II.2: Vecteur pGL3-Basic

II. Matériel et méthodes

1. Souche bactérienne et plasmides

1.1 Souche bactérienne

La souche bactérienne utilisée dans ce mémoire est *E.coli* DH11S. Son génotype est le suivant : F⁻ *mcrA* Δ (*mrr hsdRMS mcrBC*) Δ (*lac-proAB*) Δ (*recA1398*)*deoR supE rpsL srl thi*/F' *proAB*⁺ *lacI*^qZ Δ M15.

Ce sont des cellules qui peuvent être transformées par des phagémides, et dont l'efficacité de transformation est supérieure à 10⁸ transformants/ μ g. La souche DH11S possède des caractéristiques permettant une efficacité de transformation optimale. Le type *recA1398* (qui accroît la stabilité des inserts) ou la suppression du système dégradant l'ADN contenant des cytosines méthylées et des résidus adénine, en sont des exemples.

1.2 Plasmides utilisés

1.2.1 pBluescript

Ce vecteur est un phagémide, c'est-à-dire un hybride entre un plasmide et un phage, de 2958 bp (FIG.II.1). Il porte une origine de répllication pour *E.coli* et l'origine de répllication phagienne f 1. Afin de sélectionner des clones recombinants, il porte également le gène codant la résistance à l'ampicilline, ainsi que le gène *LacZ* (possédant à son extrémité 5', un promoteur inductible permettant l'expression d'une protéine de fusion, grâce au produit du gène de la β -galactosidase), permettant par α -complémentation, le test bleu/blanc. Enfin, il détient un site multiple de clonage à l'extrémité N-terminale du gène *LacZ*. Ces sites n'altèrent pas la complémentation mais si un fragment est cloné dans l'un d'entre eux, le gène est disrupté et la complémentation est alors impossible.

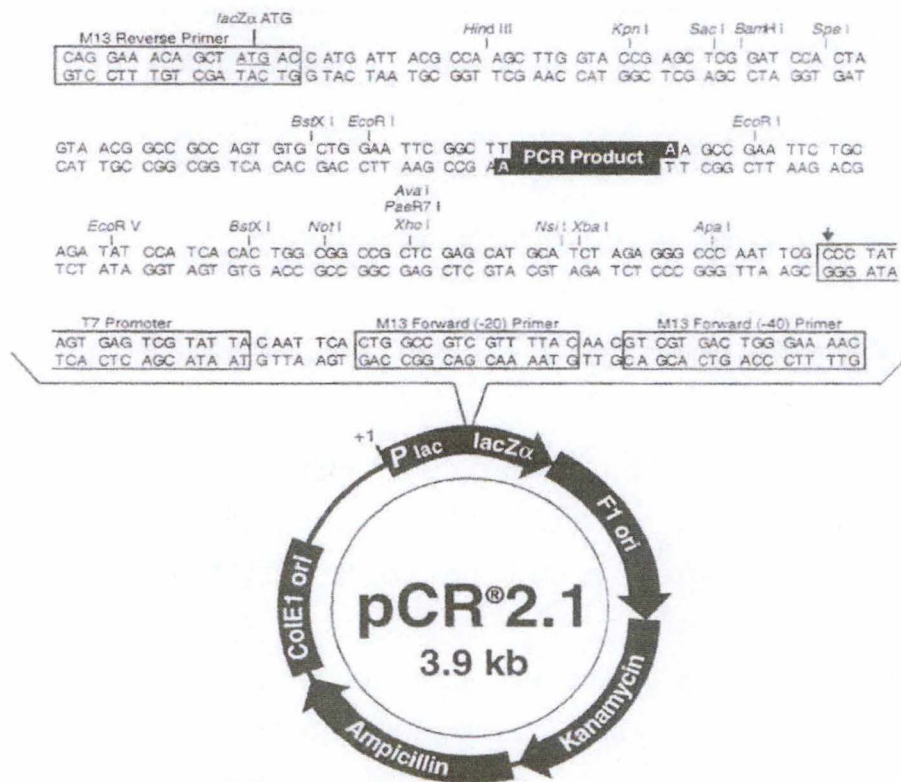


Fig.II.3:Vecteur pCR2.1

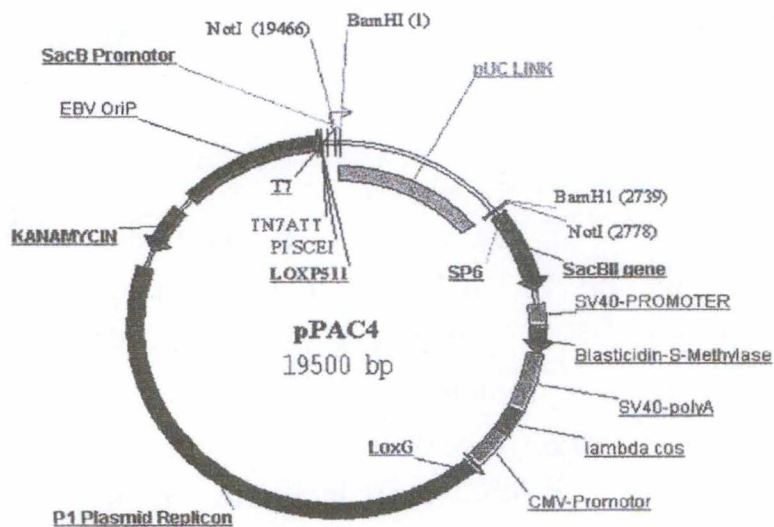


Fig.II.4: Vecteur pPAC4

1.2.2 pGL3-Basic

Le vecteur pGL3-Basic est un phagémide de 4818bp (Fig.II.2). Il porte une origine de réplication pour *E.coli* et l'origine de réplication phagienne f 1, un site multiple de clonage, ainsi que le gène de résistance à l'ampicilline. Il détient également un signal poly (A) lié à un site de pause transcriptionnel, situé en amont du site multiple de clonage, permettant de terminer les mauvais transcrits, réduisant ainsi le bruit de fond. Enfin, un signal poly (A) SV40, localisé à l'extrémité 3' de l'ADNc codant la Luciférase *firefly*, permet une terminaison de la transcription efficace ainsi que la polyadénylation de l'ARNm.

1.2.3 pCR2.1

Ce vecteur est également un phagémide de 3900bp (Fig.II.3). Il porte le gène de résistance à l'ampicilline et à la kanamycine. Il contient aussi une origine de réplication pour *E.coli* et l'origine de réplication phagienne f 1. Enfin, il porte le gène LacZ flanqué en 5' de son promoteur inductible.

1.2.4 pPAC4

Les PACs (P-1 derived artificial chromosome) utilisent un vecteur dérivé du phage P1 (Fig.II.4). Le vecteur utilisé, d'une taille de 19500bp, porte entre autre :

- l'origine de réplication du bactériophage P1,
- le gène de résistance à la kanamycine,
- le site de clonage *BamH1*, flanqué des séquences promotrices T7 et SP6,
- deux sites de restriction *Not1* immédiatement flanqué du segment T7-*BamH1*-SP6,

2. Préparation d'ADN plasmidique

2.1 Clones provenant d'une banque génomique PAC

Les clones sont étalés sur des boîtes LB + kanamycine (25 mg/ml). Celles-ci sont incubées une nuit à 37°C. Une colonie est prélevée et mise en culture dans un milieu contenant 5 ml de LB et de la kanamycine. Suit une incubation à 37°C pendant une nuit.

Cette culture est diluée 30X dans un milieu identique à celui utilisé précédemment et incubé 2 heures à 37°C. Bien que le vecteur n'est présent qu'à raison d'une seule copie par cellule, il est possible d'augmenter le nombre de copies par addition d'IPTG (concentration finale : 1mM), activateur de l'opéron lytique multicopie du phage P1. La culture est ensuite incubée durant 4 heures à 37°C.

L'ADN est ensuite récupéré par lyse alcaline à l'aide du kit Nucleobond AX500 (Macherey-Nagel).

2.2 Clones provenant de phagémide

Les clones sont étalés sur des boîtes LB + ampicilline (100mg/ml) et incubés durant une nuit à 37°C.

2.2.1 Midipréparation d'ADN

Une colonie est mise en culture dans 5 ml de LBA. Une incubation à 37°C durant la nuit est alors réalisée. La culture est ensuite diluée 30X dans un milieu LBA et est incubée à 37°C durant 12 à 16 heures. L'ADN est alors récupéré par lyse alcaline à l'aide du kit Qiagen Plasmid Midi (Qiagen S.A.).

2.2.2 Minipréparation d'ADN

Une colonie est placée dans 10 ml de milieu LBA. Une incubation pendant une nuit à 37°C, est requise.

Pour 1.5 ml de culture, une centrifugation d'une durée de 30 secondes permet de récupérer les clones. Ces derniers sont repris, respectivement, dans ;

- .100 μ l de GTE (50 mM glucose, 25 mM Tris-HCl pH8, 10 mM EDTA), sont vortexés et maintenus durant 5' à température ambiante,
- .150 μ l de NaOH/SDS (0,2 N NaOH et SDS 1%) permettant la lyse des cellules, incubation 5' dans de la glace,
- .120 μ l de KAc/HAc (12 ml KAc 5 M + 2,3 ml HA + 5,7 ml H₂O) neutralisant le lysat.

Après avoir été mélangé, la solution est centrifugée pendant 5 minutes. Une extraction au phénol-chloroforme est alors réalisée à partir du surnageant. Ce dernier est repris dans un volume phénol et un volume chloroforme. Après 5 minutes de centrifugation, la phase aqueuse est reprise dans un volume chloroforme et également centrifugée durant 5 minutes.

La phase aqueuse est lavée avec deux volumes d'éthanol 100 % et centrifugée pendant 10 minutes. Après avoir décanté, rincé avec de l'éthanol 70 % et séché le culot sous vide, ce dernier est repris dans 10 μ l d'H₂O ou de TE 1X. Il est alors traité à la RNase durant 60 minutes à 37°C, éliminant ainsi, toute trace d'ARN.

3.Quantification d'ADN

La concentration en ADN d'un échantillon est déterminée à l'aide d'un spectrophotomètre (GeneQuant RNA/DNA calculator).

Cet appareil mesure l'absorbance pour une longueur d'onde donnée, de l'échantillon d'ADN. Ce dernier, placé dans une cellule adéquate, est une dilution de l'échantillon « mère ».

La formule suivante nous permet de calculer la concentration en μ g/ml ;

$$(A_{260} \times \text{facteur}) \times \text{dilution} = \text{concentration en } \mu\text{g/ml}$$

A₂₆₀ : absorbance mesurée pour une longueur d'onde de 260 nm.

Facteur : 1 DO (densité optique) pour la longueur d'onde précitée est égale à 50 $\mu\text{g/ml}$ pour l'ADN double brin.

4. Réaction de polymérisation en chaîne (ou PCR)

4.1 Conditions classiques de PCR

Les réactions de PCR sont réalisées à partir de 20 ng d'ADN de PAC ou de 200 ng d'ADN génomique comme matrice. Les oligonucléotides utilisés sont synthétisés par la firme Eurogentec.

La concentration des différents composants du mélange réactionnel est la suivante :

- .ADN 20 ng ou 200 ng
- .4dNTPs 0,2 mM
- .Tampon 75 mM Tris-HCl(pH9), 2 mM MgCl_2 , 20 mM $(\text{NH}_4)_2 \text{SO}_4$, 50 mM KCl
- .Bétaine 1 M
- .DMSO 5 %
- .Amorces 0,2 μM
- .Enzyme 0,5 U pour 20 μL de réaction

L'enzyme utilisée est la DNA polymérase thermostable Tth (*Thermus thermophilus*) de la firme Biotools.

Les cycles sont réalisés dans un appareil MJ Research dans les conditions suivantes ;

dénaturation initiale à 95°C durant 5', suivi de 30 à 35 cycles comprenant chacun une phase de dénaturation à 95°C pendant 30'', une phase d'hybridation à 55-65°C durant 30'', ainsi qu'une phase d'élongation à 72°C pendant 45''.

Les 30 cycles sont suivis d'une phase d'élongation de 72°C durant 10'.

La température d'hybridation est adaptée en fonction des amorces utilisées. En général, la température d'hybridation est de 5°C inférieure au T_m de l'oligonucléotide.

L'équation ci-dessous nous permet de déterminer le T_m pour des oligos d'environ 20 bases :

$$T_m = 2^{\circ}\text{C}(\text{A}+\text{T})+4^{\circ}\text{C}(\text{G}+\text{C})$$

4.2 Conditions particulières de PCR

PCR longue

L'amplification de fragments de taille supérieure à 3 kb nécessite l'utilisation d'ADN polymérase possédant une activité de correction d'épreuve (« proofreading activity », c'est-à-dire une activité 3'-5' exonucléase).

Nous avons utilisé l'Elongase (Life Technologies), composé d'un mélange d'ADN polymérase de *Thermus aquaticus* et de *Pyrococcus*.

Le mélange réactionnel est constitué de deux « mix ». La concentration des composants du premier mix est la suivante ;

.ADN 100 ng
.4dNTPs 0,2 mM
.Amorces 0,2 μM
.DMSO 5 %

La concentration des constituants du second mix est la suivante ;

.TamponB 300 mM Tris- SO_4 , (pH 9,1), 90 mM $(\text{NH}_4)_2\text{SO}_4$ et 10 mM MgSO_4
.Enzyme 1U pour 15 μl de réaction

Les cycles sont réalisés dans un appareil MJ Research dans les conditions suivantes ;

dénaturation initiale à 94°C durant 30'', suivi de 35 cycles comprenant chacun une phase de dénaturation à 94°C pendant 30'', une phase d'hybridation variant de 55 à 65°C (suivant le T_m des amorces) durant 30'', ainsi que d'une phase d'élongation à 68°C pendant une durée de 45-60'' par kb.

5.Électrophorèse d'ADN sur gel d'agarose

Cette technique permet de séparer des fragments d'ADN, et d'en déterminer la taille. En effet, la distance parcourue dans le gel par une molécule est inversement proportionnelle au logarithme de sa longueur. Le degré de réticulation du gel varie en fonction de la concentration en agarose, laquelle est adaptée en fonction de la taille des fragments à séparer. Le pourcentage des gels utilisés varie entre 0,8 et 2 %. La taille des fragments est déterminée par l'utilisation de marqueurs de poids moléculaire calibrés (Life Technologies). Les fragments d'ADN sont visualisés grâce à du bromure d'éthidium. Celui-ci se lie à l'ADN bicaténaire en s'intercalant entre les bases de la double hélice. Etant fluorescente, la molécule marque l'ADN et celui-ci est donc visible sous illumination UV. Il faut cependant noter que la fluorescence d'un fragment d'ADN est proportionnelle à sa taille, mais aussi à la quantité d'ADN présent. Ainsi, en connaissant la quantité d'ADN du marqueur de taille, on peut estimer la quantité d'ADN du fragment par comparaison de l'intensité de la fluorescence.

Le gel est réalisé de la manière suivante ;

- .porter à ébullition un volume approprié de tampon TBE 1X additionné d'agarose (Boehringer), jusqu'à obtention d'une solution limpide,
- .lorsque la solution est environ à 60°C, ajouter 2 μ l de BrEth (2mg/ml) et couler l'agarose sur le support *ad hoc* muni d'un peigne,
- .après solidification du gel, l'immerger dans une cuve d'électrophorèse contenant du TBE 1X.

Les échantillons sont déposés dans les puits après ajout de la solution stop (20 % du volume à déposer) constituée de ;

- .sucrose 50 %,
- .EDTA 100 mM,
- .SDS 1 %,
- .bleu de bromophénol 0.1%.

La migration peut, alors, avoir lieu au voltage désiré. L'ADN est visualisé sous UV (transilluminateur, longueur d'onde à 320nm).

6. Purification des produits PCR en vue du séquençage

Le produit PCR est purifié à l'aide de colonnes Wizard (kit Wizard PCR Preps DNA Purification System de la firme Promega). La phase aqueuse est placée dans un Eppendorf dans lequel on ajoute 1ml de résine. Après avoir vortexé, la solution est déposée sur la colonne à l'aide d'une seringue. Ensuite, la colonne est lavée à deux reprises avec 1ml d'isopropanol 80 % et centrifugée durant 30 secondes. L'ADN est alors élué dans 50 μ l de TE 1X (10mM Tris, 1mM EDTA). Pour les fragments de tailles supérieures à 3kb, il est possible d'augmenter le pourcentage de récupération en ajoutant du TE 1X chaud (65-80°C).

Si le produit PCR contient plusieurs bandes, le fragment d'intérêt est récupéré de la manière suivante ; le produit d'amplification est déposé sur gel d'agarose « SeaPlaque GTG » (de la firme FMC). Après migration, le fragment d'intérêt est prélevé et placé dans un Eppendorf. La purification est alors identique à celle décrite ci-dessus.

7. Séquençage

La réaction de séquençage, basée sur la méthode de terminaison de chaînes par incorporation de didéoxynucléotides, consiste en une amplification linéaire du produit à partir d'une amorce spécifique, au moyen d'une ADN polymérase thermostable et d'un mélange de déoxy et de didéoxynucléotides qui induisent des terminaisons aléatoires.

Le séquençage automatique (séquenceur ABI PRISM 377 de Perkin Elmer) est basé sur l'utilisation de didéoxynucléotides chacun couplé à une molécule fluorescente différente. La taille des molécules détectées et leur longueur d'onde de fluorescence permettent de reconstituer la séquence du fragment analysé.

7.1 Séquençage sur de l'ADN de phagémide

La réaction de séquençage est réalisée à l'aide du kit Big Dye Terminator FS (Perkin Helmer), qui contient un pré-mélange des réactifs suivants : les 4 terminateurs ddATP, ddCTP, ddGTP, ddTTP couplés à un fluorochrome, les 4 déoxynucléotides dATP, dTTP, dGTP, dCTP et de l'ADN polymérase AmpliTaq (volume utilisé : 2 μ l). Cette réaction nécessite, en outre, entre 40 et 100 ng de produit purifié (en fonction de la taille des produits PCR), 0,1 μ M d'amorce, du DMSO 5 % et du tampon Tris-HCl MgCl₂, dans un volume final de 20 μ l.

Les cycles sont réalisés dans un appareil MJ Research dans les conditions suivantes ;

une première phase de dénaturation durant 30'' à 96°C, suivi d'une phase d'hybridation à 50°C pendant 15'' et enfin une phase d'élongation à 60°C durant 4'. Ces phases sont répétées 25x.


Avant d'être déposé sur gel, ces produits de séquençage sont purifiés (élimination de l'excès des terminateurs) par précipitation à l'éthanol. Une solution contenant 2 μ l d'acétate de sodium 3 M pH 4,6 et 50 μ l d'éthanol 95 % est ajoutée à la phase aqueuse, vortexée et placée sur glace 10 minutes. Elle est ensuite centrifugée 30 minutes à 14000 rpm. Le surnageant est éliminé et le culot est lavé avec 250 μ l d'éthanol 70 % puis séché au speed vack. Le culot est ensuite stocké à -20°C.


7.2 Séquençage sur de l'ADN de PAC


La réaction de séquençage est également réalisée à partir du kit Big Dye Terminator FS (Perkin Helmer). Cependant, le séquençage sur ADN de PAC est légèrement différent du séquençage sur ADN de phagémide.

En effet, cette réaction requiert 1,5 μ g d'ADN, 4 μ l du pré-mélange (ou mix), 12 μ l de tampon et 10 μ M d'amorce, dans un volume final de 40 μ l.

Les cycles, au nombre de 40, sont réalisés dans les conditions suivantes ;

 Superpool 1
= plaques 1 à 8

 Superpool 2
= plaques 9 à 16

 Superpool 30
= plaques 233 à 240

Superpools

1	2	3	4

28	29	30
		X

Diagram illustrating a 3D array structure (Plaques, Lignes, Colonnes) and its corresponding 2D grid representation.

The 3D structure shows a stack of 8 Plaques (8), each containing 16 Lignes (16) and 24 Colonnes (24).

The 2D grid representation shows the mapping of Lignes (a, b, c, d, e, ..., p) and Colonnes (1, 2, 3, 4, 5, ..., 24) to Plaques (233, 234, 235, 236, 237, 238, 239, 240).

The grid shows a cross (X) in the cell corresponding to Ligne 'a' and Colonnes '3', indicating a specific data point or location within the array.

C. Dans cet exemple le clone positif est P3233

Fig.II.5: Illustration du principe de criblage d'une banque PAC par PCR selon le système de super-pools et de pools.

une phase de dénaturation à 95°C pendant 30'', suivi d'une phase d'hybridation à 50°C durant 30'' et enfin, d'une phase d'élongation à 60°C d'une durée de 4'.

Les produits de réaction de séquençage doivent être purifiés à l'aide de colonnes Centri-Sep (Perkin-Elmer). Ces colonnes sont hydratées avec 800µl d'eau, durant une heure avant leur utilisation. Une attention particulière sera portée au fait qu'il ne peut subsister aucune bulle dans la colonne.

On laisse ensuite s'écouler l'eau et les colonnes sont centrifugées 2 minutes à 3000 rpm. Le produit de la réaction de séquençage est déposé délicatement à la surface de la résine. Le tout est centrifugé 2 minutes à 3000 rpm. L'échantillon est récupéré et séché au speed-vack. Le culot peut être stocké à -20°C.

8. Criblage d'une banque d'ADN génomique

Nous avons criblé la banque RZPD N°711. Il s'agit d'une banque d'ADN génomique souris. Le vecteur pPAC4, dans lequel les fragments d'ADN sont insérés, est propagé dans *E.coli* DH10B et est présent à raison d'une copie par cellule. La taille des fragments insérés est en moyenne de 146 kb.

Le criblage de la banque est réalisé par le système de super-pools et de pools (Fig.II.5). La totalité de la banque est représentée par 84 super-pools qui correspondent chacun au regroupement des clones contenus dans 8 microplaques, soit 3072 clones. Les super-pools sont testés par 84 PCR. Lorsqu'un super-pool est positif, un second criblage est effectué afin de déterminer le clone positif qu'il contient. Pour identifier ce dernier, le second criblage doit donner trois signaux positifs correspondant à une ligne, une plaque et une colonne.

Ce système de criblage permet de réduire considérablement le nombre de PCR à effectuer pour identifier un clone positif. En effet, il suffit d'une réaction PCR pour chacun des 84 super-pools ainsi que de 48 réactions pour le second criblage (16 lignes, 24 colonnes et 8 microplaques). En tout, 78 réactions sont donc nécessaires pour identifier un clone positif parmi les 258 048 clones constituant la banque.

9. Electrophorèse d'ADN dans un champ électrique pulsé

Cette technique permet de séparer des fragments d'ADN dont la taille est supérieure à 50 kb. Les fragments d'ADN, soumis à un champ électrique alternatif, sont capables de se réorienter dans le gel d'agarose. Nous avons utilisé l'appareil CHEF-DR II (Bio-Rad) qui produit un champ électrique homogène à partir d'un dispositif hexagonal de 24 électrodes. Celles-ci sont activées de manière à générer en alternance deux champs électriques dont les vecteurs de courant forment un angle de 120° l'un par rapport à l'autre. La concentration en agarose, la concentration et la température du tampon, le voltage, le temps de pulse et la durée totale de l'électrophorèse sont les facteurs affectant l'électrophorèse.

La taille des fragments d'ADN est déterminée par l'utilisation de marqueurs de poids moléculaire calibrés (Sigma).

Nous avons utilisé le programme suivant ;

- .durée d'impulsion : 0,2-22 s
- .durée de la migration : 15 h
- .voltage : 200 V
- .gel : agarose SeaKem (FMC)
- .degré de réticulation : 0,8 %
- .tampon : TBE 0,5X
- .température du tampon : 14°C

Après avoir migré, le gel est révélé au bromure d'éthidium et la visualisation s'effectue aux UV.

10. Extraction d'ADN génomique de souris

Si l'on veut travailler à partir du foie, les animaux doivent être à jeun depuis 24 heures. Immédiatement après la mise à mort, effectuée par dislocation cervicale, le corps est généreusement aspergé d'alcool dénaturé.

Une fois prélevé, le tissu est transféré dans un homogénéiseur de Dounce avec 1 ou 2 gouttes d'eau distillée et broyé. 5 ml de solution d'homogénéisation

(Tris-HCl 10 mM à pH 8, EDTA 10 mM à pH 8, SDS 1 %, protéinase K 100 µg/ml) sont ajoutés et l'homogénéisation est poursuivie. Ce mélange est transféré dans un tube de 50 ml. Le Dounce est rincé avec 5 ml de solution supplémentaire. Les 10 ml d'homogénats sont incubés pendant toute une nuit à 55°C, sous agitation. Le lendemain, l'homogénat est traité à la RNase A, ajoutée à la concentration de 100 µg/ml, à 37°C durant 1 heure.

L'ADN est extrait au phénol/chloroforme. Deux volumes d'éthanol sont ajoutés à la phase aqueuse et l'ADN précipité est récupéré avec une baguette en verre. L'ADN est ensuite rincé à l'éthanol 70 %, séché et resuspendu dans du TE 1X sous agitation modérée pendant au moins 24 heures. La concentration en ADN est estimée par détermination de l'absorbance à 260nm.

11. Southern Blot et marquage radioactif

Les techniques d'hybridation permettent de détecter une séquence donnée parmi une population complexe à l'aide d'une sonde d'ADN dont la séquence est complémentaire de celle recherchée et qui est marquée par incorporation d'isotopes radioactifs. Il est alors possible de déterminer si une séquence correspondant à la sonde est présente dans l'ADN cible.

L'ADN génomique est digéré par des enzymes de restriction selon les recommandations de la firme. Les fragments d'ADN sont soumis à une électrophorèse en gel d'agarose 0.8 % dans du tampon TBE 1X (0,09 M Tris-Borate, 0,002 M EDTA pH 8), et fixés sur membrane de nylon (Nylon Hybond-N⁺) par transfert alcalin (Southern blotting) selon le protocole d'Amersham.

Les marquages radioactifs de fragments d'ADN au phosphore-32 sont réalisés par incorporation de (α^{32} P)dCTP (activité spécifique de 3000 Ci/mM ; ICN) à l'aide du kit de random priming RadPrime DNA Labeling System (Life Technologies). Habituellement, 25 ng d'ADN sont marqués avec 5 µl de (α^{32} P)dCTP dans un volume final de 50 µl. Les sondes sont séparées des nucléotides non incorporés par passage sur une spin colonne de résine Ultrogel AcA54.

La pré-hybridation est réalisée à 65°C dans du SSPE 1,5X, SDS 1 %, Blotto 0,5 % et ADN carrier 0,5 mg/ml. L'hybridation se déroule la nuit, puis la membrane est lavée (3 ou 4 fois) avec une solution contenant du SDS 0,1 % et du SSPE.

Enfin, la membrane est séchée et exposée à un film Ray Max Universal (ICN) durant une période pouvant aller de 1 à 3 jours à -70°C .

12. Constructions génétiques et transformation de cellules

12.1 Constructions génétiques

12.1.1 Restriction de l'ADN

Le clivage de l'ADN bicaténaire à des sites précis se fait par des endonucléases qui reconnaissent de courtes séquences palindromiques. En pratique, la réaction se fait par incubation de l'ADN avec une ou des enzymes de restriction.

Mélange de restriction :

- . X μg d'ADN,
- . tampon de restriction 10 X : 1/10 du volume final (tampon commercial fourni avec l'enzyme),
- . enzyme de restriction : au maximum 1/10 du volume final,
- . eau distillée stérile pour atteindre le volume final.

Une incubation de 2 heures à la température optimale pour l'enzyme (selon les recommandations de la firme) est ensuite réalisée.

12.1.2 Déphosphorylation

Après une restriction d'un plasmide, les extrémités 5' peuvent être déphosphorylées afin d'augmenter l'efficacité de clonages futurs en empêchant le plasmide de se refermer sur lui-même. Cette réaction de déphosphorylation est effectuée par la CIP (Calf Intestine Phosphatase, Biolabs). Elle catalyse l'hydrolyse des résidus 5'-phosphate de l'ADN. Le produit déphosphorylé possède ainsi une extrémité 5'-hydroxyl.

Mélange réactionnel :

- . ADN restreint,
- . CIP 1 U/ml
- . tampon de CIP 10 X
- . porter à volume avec de l'eau stérile

Après une incubation de 2 heures à 37°C, la réaction de déphosphorylation est stoppée en incubant le mélange réactionnel à 72°C durant 10 minutes.

12.1.3 Réaction de ligation

Pour lier de manière covalente deux fragments d'ADN, on utilise l'enzyme ADN ligase du phage T4. Elle catalyse la formation de liens phosphodiester entre l'hydroxyl 3' d'une extrémité d'ADN et le phosphate 5' d'une extrémité proche. L'enzyme pour être active, nécessite la présence d'ATP et d'ions Mg^{++} .

Le mélange réactionnel se compose comme suit ;

- . les deux fragments d'ADN à liguer,
- . 1 μ l d'ADN ligase T4, 1 U/ml (Biolabs),
- . tampon de ligation 10X (Biolabs),
- . mise à volume avec de l'eau distillée stérile.

Le mélange ainsi préparé est incubé à 14°C durant 10 à 14 heures pour des extrémités cohésives et quelques heures à environ 25°C pour des bouts francs. La « ligation » peut être conservée à - 20°C.

Un vecteur restreint, dont le mélange de ligation ne contient pas de ligase, peut être utilisé comme témoin. En effet, étant linéaire, le vecteur ne peut transformer efficacement les *E.coli* compétentes. Les colonies apparues sur milieu sélectif après transformation permettent d'évaluer la proportion des transformants dus aux vecteurs non restreints.

12.1.4 Réaction de « fill in » et de rognage

Pour passer d'une extrémité cohésive d'ADN (issue d'une restriction enzymatique) à une extrémité à bouts francs, on utilise l'ADN polymérase I large fragment (Klenow, de la firme Biolabs).

Cette polymérase possède une activité 3'-5' exonucléase élevée mais est dépourvue d'activité 5'-3' exonucléase. Les extrémités 5' protrudantes sont complétées et les extrémités 3' protrudantes sont rognées.

Le mélange réactionnel est le suivant ;

- . x μ g d'ADN restreint,
- . tampon de polymérase 10X,
- . solution 1 mM en dATP, dGTP, dCTP, dTTP,
- . ADN polymérase (1 U/ μ g)

Une incubation de 15 minutes à 25°C est requise. L'inactivation de l'enzyme s'effectue à 75°C durant 10 minutes.

12.2 Transformation bactérienne

12.2.1 Transformation bactérienne avec les constructions génétiques

La transformation de bactéries compétentes consiste à faire entrer de l'ADN exogène (plasmide) dans le microorganisme.

À cette fin, la paroi et la membrane d'*E.coli* doivent être perméabilisées à l'ADN : la bactérie est alors dite compétente. Après transformation par électroporation ou choc thermique, permettant au complexe de rentrer dans la cellule, un temps d'incubation en milieu riche non sélectif permet aux bactéries de régénérer une paroi normale, et aux gènes de résistance introduits de s'exprimer. Les transformants sont isolés par culture sur boîtes de milieu sélectif contenant l'antibiotique dont le gène de résistance est apporté par le plasmide.

La méthode de transformation utilisée est l'électroporation, où la survie et la transformation bactérienne dépendent de l'intensité et de la durée du choc électrique.

La préparation des cellules électrocompétentes est la suivante ;

- . inoculer 5 ml de LB liquide avec une culture fraîche de bactéries *E.coli.*, agiter pendant 5 h à 37°C,
- . inoculer 500 ml de culture avec la préculture. Agiter à 37°C et laisser pousser à une DO de 0.5 à 0.7, mesurée à 550 nm,
- . placer la culture 15 à 30 minutes dans la glace, centrifuger à 4°C à 4200 rpm pendant 15 minutes,
- . éliminer le surnageant. Resuspendre le culot dans 500 ml d'eau distillée stérile (d.s.) froide. Centrifuger comme précédemment,
- . resuspendre dans 500 ml d'eau d.s. froide et centrifuger,
- . resuspendre dans 40 ml d'eau d.s. froide contenant 10% de glycérol. Centrifuger,
- . resuspendre dans un volume final de 2 à 3 ml d'eau d.s. froide contenant 10 % de glycérol. La concentration cellulaire est d'environ $1 \text{ à } 3.10^{10}$ cellules/ml. Cette suspension peut être aliquotée sur glace et conservée à -80°C (pendant environ 6 mois).

L'électroporation se déroule comme suit ;

- . laisser dégeler 100 ml de cellules électrocompétentes sur glace,
- . y ajouter entre 10 et 100 ng d'ADN plasmidique ,
- . mélanger et placer la solution 1 minute sur glace,
- . transférer sans faire de bulles, le contenu de l'eppendorf, dans la cuvette d'électroporation (se trouvant dans la glace)
- . placer la cuvette dans l'électroporateur (Bio Rad Gene Pulser™),
- . régler l'appareil sur : 2350 ou 2400 Volts
- . réaliser le choc électrique,
- . ajouter rapidement 1ml de LB sans antibiotique,
- . transférer dans un eppendorf et laisser 1 heure à 37°C,
- . étaler sur milieu sélectif (100 à 250 ml/boîte) et incubé une nuit à 37°C.

12.2.2 Transformation bactérienne avec un produit PCR cloné à l'aide du kit TA cloning (Invitrogen)

Principe du kit : de nombreuses *Taq* polymérases ajoutent une déoxyadénosine (A) à l'extrémité 3' du produit PCR. Le vecteur linéarisé fourni dans le kit possède une déoxithymidine à son extrémité 3', ce qui permet au produit PCR de se lier de manière efficace.

La méthode comprend trois étapes. La première étape consiste à cloner le produit PCR dans le vecteur pCR2.1. Un mélange de 5 μ l d'eau stérile, 1 μ l de tampon de ligation, 2 μ l de vecteur pCR2.1, 1 μ l de produit PCR et 1 μ l de T4 DNA ligase est incubé une nuit à 14°C.

La seconde étape consiste à transformer des cellules compétentes. Les cellules (One Shot Competent Cells) sont décongelées sur glace, puis 2 μ l de β -mercaptoéthanol 0.5 M et 1 à 2 μ l de la réaction de ligation sont ajoutés. Un choc thermique à 42°C est réalisé pendant 30 secondes, puis les cellules sont placées sur glace durant 2 minutes. Après addition de 450 μ l de milieu SOC (2 % tryptone, 0.5 % d'extrait de levure, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2 \cdot 6 H_2O$, 20 mM glucose), les bactéries sont placées 1 heure à 37°C. Enfin, après cette période de récupération, 50 μ l et 200 μ l de chaque transformation sont étalés sur des boîtes LB contenant 50 μ g/ml d'ampicilline, de l'inducteur IPTG et le substrat X-gal (50 mg/ml). Les boîtes sont incubées à 37°C pendant au moins 18 heures. Les colonies blanches, dans lesquelles le gène de la β -galactosidase a été inactivé par la présence de l'insert, sont prélevées et analysées par PCR afin de s'assurer qu'elles contiennent bien l'insert.

Oligonucléotides spécifiques du vecteur pCR2.1 :

M13F : 5' CGC CAG GGT TTT CCC AGT CAC GAC 3'

M13R : 5' TCA CAC AGG AAA CAG CTA TGA C 3'

13. Mise en collection de bactéries

Une culture de 5 ml d'une colonie isolée, est incubée 12 heures à 37°C, dans un milieu contenant du LB ainsi qu'un antibiotique approprié.

500 µl de cette culture sont prélevés et mélangés à 500 µl de LB-40 % glycérol. Le mélange est ensuite stocké à -70°C.

14. Transfection

14.1 Les plasmides

14.1.1 Le plasmide test

Le plasmide pGL3-Basic a été décrit antérieurement au point 1.2.2.

14.1.2 Le plasmide contrôle

Le plasmide utilisé est pRL-TK. Ce vecteur contrôle contient l'ADNc codant la *Renilla* luciférase (*Rluc*) cloné à partir d'un anthozoaire coelentéré *Renilla reniformis*, un organisme marin.

Qui plus est, il porte le promoteur HSV-TK (herpes simplex virus thymidine kinase) à l'extrémité 5' de *Rluc*. Il permet une expression constitutive dans les cellules de tissus embryonnaire et mature de mammifères.

14.2 Passage des cellules

La lignée cellulaire utilisée est appelée P19. Elle provient d'un teratocarcinome de souris C3H/He, produite par greffage d'un embryon de 7 jours, dans le testicule d'une souris mâle adulte.

Nous avons procédé de la manière suivante ;

. les cellules sont cultivées dans du milieu DMEM (Life Technologies) auquel on a ajouté 7.5 % de sérum bovin et 2.5 % de sérum fœtal bovin (Life Technologies),

- . les cellules sont passées toutes les 48 heures, à une dilution 1/10 selon le protocole suivant :
- . laver avec 5 ml de PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1.4 mM KH_2PO_4),
- . ajouter 1 ml de trypsine-EDTA (1 mM EDTA, 0.025 % trypsine dans du PBS ; Sigma),
- . incuber 5 minutes à 37°C (les cellules doivent facilement se détacher de la surface du tube),
- . ajouter 5 ml de milieu DMEM + sérum (le sérum inactive la trypsine) et centrifuger 5 minutes à 1000 rpm,
- . décanter et reprendre le culot dans 5 ml de ce même milieu.
- . transférer 500 μl dans une boîte de culture contenant 5 ml de milieu.

14.3 Transfection par la Lipofectamine

De nombreuses méthodes ont été développées pour transférer l'ADN à l'intérieur des cellules eucaryotes lors d'études concernant la régulation de l'expression des gènes. Ces méthodes incluent l'utilisation de phosphate de calcium, des polycations, des liposomes, des rétrovirus, des micro-injections et des électroporations.

Nous avons utilisé la Lipofectamine2000 (Life Technologies). Ce produit est un agent lipofectant chargé positivement, qui permet de masquer les charges négatives de l'ADN plasmidique. Cet agent interagit avec les membranes cellulaires de manière à délivrer directement le matériel plasmidique dans le cytoplasme.

Nous avons réalisé la transfection de la manière suivante ;

- . 900 ng d'ADN à transférer (solubilisé dans du TE 1X) dans un tube eppendorf,
- . ajout d'un mix contenant 50 μl d'OptiMem (Life Technologies) et 100 ng de pRL-TK,
- . après 5 minutes à température ambiante, ajout d'un second mix constitué de 50 μl d'OptiMem et de 1 μl de Lipofectamine2000.

. incuber 20 minutes à température ambiante, et déposer la solution dans le puit adéquat. Après une incubation à 37°C de 24 à 48 heures, la lyse cellulaire peut être effectuée.

Nous avons effectué nos transfections dans une plaque multipuits de 24 puits, avec 10⁵ cellules par puit.

14.4 Préparation du lysat cellulaire

Après avoir enlevé le milieu de culture, 200 µl de tampon PBS 1X sont ajoutés dans chaque puit. Ensuite, il est enlevé et 100 µl d'une solution de lyse (PLB :Passive Lysis Buffer 1X) est ajoutée.

Après incubation, sous agitation, d'une durée de 15 minutes à température ambiante, le lysat est transféré dans un eppendorf. Lequel est maintenu sur glace.

14.5 Méthode de détection : le système Dual-Luciferase

Ce système permet à la fois de quantifier et de normaliser les expériences de transfections. En effet, le gros problème de cette technique est d'obtenir une efficacité de transfection de culture identique. C'est pourquoi, nous sommes obligés de faire une double transfection avec d'une part le plasmide test, et d'autre part, le plasmide contrôle dont la régulation ne varie pas.

Ces deux types de plasmides possèdent un gène luciférase provenant de deux organismes différents. Les deux luciférases ont des structures différentes, et interagissent avec des substrats différents pour leur réaction en bioluminescence. Nous mesurons donc deux activités luciférases séquentiellement à partir d'un même échantillon : la *Firefly* luciférase et la *Renilla* luciférase.

L'activité de la première est mesurée après ajout du substrat LAR II.

En effet, après clivage du substrat, une émission de lumière pourra être mesurée. Ensuite, on ajoute le second substrat (Stop & Glo) qui inhibe la *Firefly* Luciférase et l'on mesure le signal lumineux provenant de la réaction enzymatique de la *Renilla* Luciférase.

A cette fin, nous avons procédé de cette manière ;

- . ajouter 10 μ l de lysat cellulaire à 100 μ l de Luciferase Assay Reagent II,
- . après avoir mélangé, effectuer la première lecture au luminomètre (Biocunter, M2010, Lumac),
- . ajouter 100 μ l de Stop & Glo et effectuer la seconde mesure.

15. Analyse informatique

La gestion et l'alignement des séquences obtenues sont réalisés à l'aide du programme Mac Vector 6.5 (Oxford Molecular). Les séquences sont également comparées à celles des bases de données grâce au logiciel BLAST, accessible via Internet. BLASTN compare une séquence d'ADN à des banques de séquences d'ADN (EMBL, Genbank, Databank, EPD), (Altschul et al., 1990).

La recherche d'îlots CpG peut s'effectuer à l'aide de programmes tels que Mac Vector 6.5 (Oxford Molecular), WebGene et GRAIL (Pedersen et al., 1999).

Plusieurs méthodes sont disponibles sur Internet pour détecter des motifs de liaison de protéines à des séquences d'ADN. Les programmes MatInspector et TFSearch reposent sur des banques de matrices (« weight matrix ») issues de TRANSFAC, la banque des facteurs de transcription eucaryotes et de leurs sites de fixation sur l'ADN (Wingender et al., 1996). Le résultat de l'analyse d'une séquence d'ADN par ces programmes consiste en une liste des sites potentiels de liaison (avec un score, suggérant la probabilité d'un site de liaison sur l'ADN) et leurs positions dans la séquence.

Résultats

-760 ctgactggaaagggctcgcgcgcgcgcgccaccttcgcgg -721

-720 ttggggccactgacgggaggtccccagacaagcgctgcccgctggggactggggggcgcgcgcccccgaggggaggggtt -641

-640 gcagccggtgtagacaaagaacagcgcggtccccgcgccttcggtttcctcgcggtcctcgagaaacgggcagccgtga -561

-560 gccgggggttcgggttccccggggcagcccgctcggtcttgcccgctttcgttcgcgcgagcgccgagcagccctgcccttc -481

-480 cgggcggagcgccccgtgggcgggagggcgcgcggtgtgttcggggcacagcccgaggtgcgggcaagcccgagggggcg -401

-400 ggacgggcgcggggggcgggggcgcggtgggacacgtgtggcgcgcgggggagggggcgcgccccggggctttaagaa -321

-320 ggtgcggagggggcgggcactttcccaagcccgccgagggcggtcgcggtgcacaccggcgcgcgcgcggtcgaggcg -241

-240 gacgacgcgctctcggcgcccgcggccccggttccccccgcgctctcgctccggcgggcccaaagtaacttcgggagcctc -161

-160 ggtctcccgcctaacttccccccgcgggctcggttgcccggaccgcgctcggtcgagcccgccgcccggctcgcccttccccg -81

-80 cacgcggctcctccgtgccggtgcctccgaaagtggatgagagagcgcgcgggggcgcgcgcgccacggagcgcgggcggc -1

1	ATG	GAG	CGC	GGC	TGC	TGG	GCG	CCG	CGG	GCT	CTC	GTC	CTG	GCC	GTG	CTG	CTG	CTG	GCG	60
1	M	E	R	G	C	W	A	P	R	A	L	V	L	A	V	L	L	L	A	20

61	ACG	CTG	AGG	GCG	CGC	GCG	GCC	ACC	GGC	TAC	TAC	CCG	CGC	TTC	TCG	CCT	TTC	TTT	TTC	CTG	120
21	T	L	R	A	R	A	A	T	G	Y	Y	P	R	F	S	P	F	F	F	L	40

121	TGC	ACC	CAC	CAC	GGG	GAG	CTG	GAA	GGG	GAT	GGG	GAG	CAG	GGC	GAG	GTG	CTC	ATT	TCC	CTG	180
41	C	T	H	H	G	E	L	E	G	D	G	E	Q	G	E	V	L	I	S	L	60

181	CAC	ATT	GCG	GGC	AAC	CCC	ACC	TAC	TAC	GTA	CCG	GGA	CAG	GAA	TAC	CAT	GTT			231
61	H	I	A	G	N	P	T	Y	Y	V	P	G	Q	E	Y	H	V			77

Fig.III.1: Séquence de la région promotrice du gène *reelin* à partir de laquelle nous avons travaillé (Royaux et al., 1997). Les bases de couleur rouge indiquent la position des sites d'initiation de la transcription

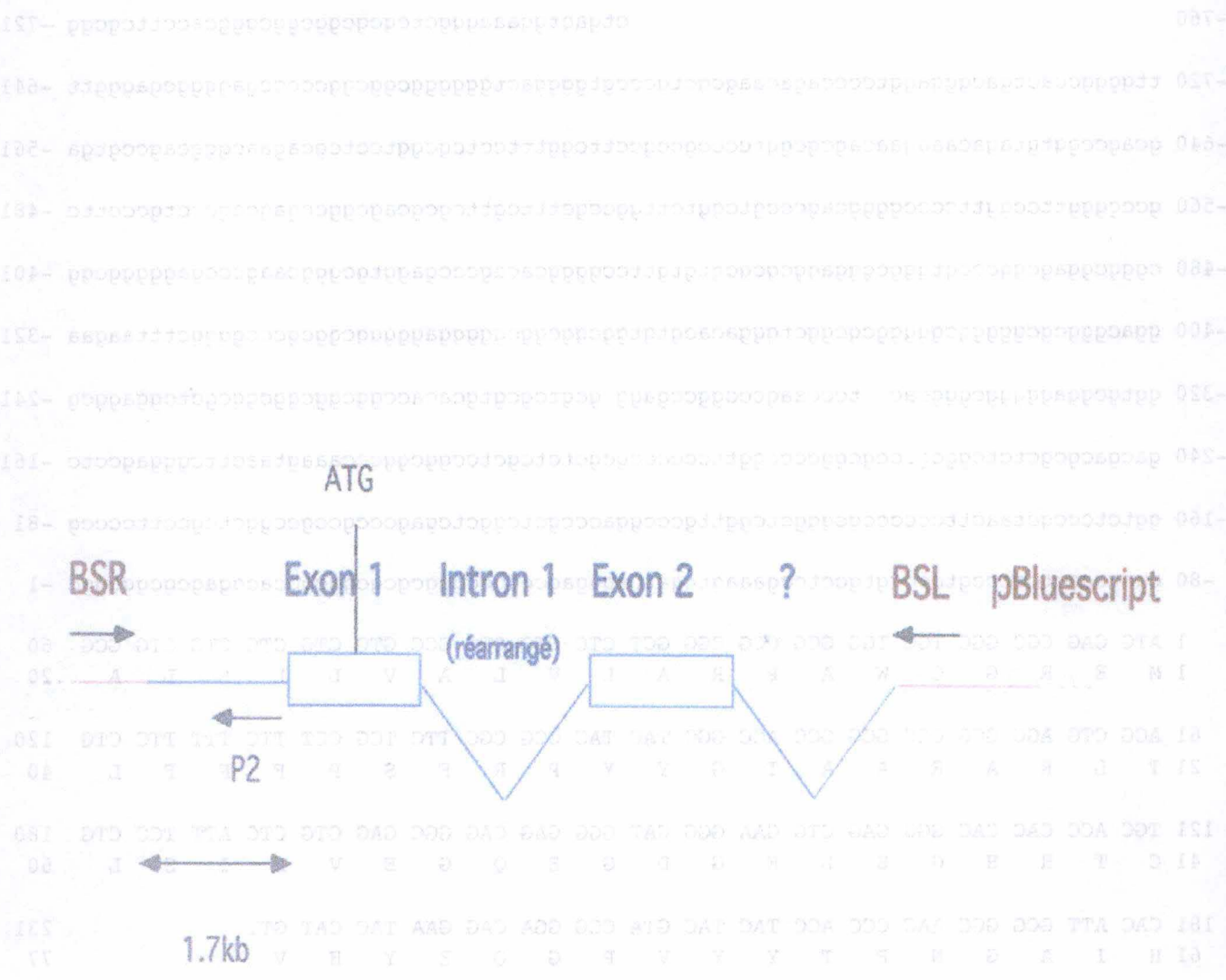


Fig.III.2: Clone génomique λ 3-7 (en bleu) sous- cloné dans pBluescript (en rouge) (Royaux et al., 1997). BSR et BSL sont les amorces spécifiques du vecteur. P2 est une amorce spécifique de la région promotrice. En amont de l'exon 1 se trouve la région promotrice que nous avons caractérisée (1.7kb). La région en aval de l'exon 2 n'est pas déterminée.

III. Résultats

1. Caractérisation de la région promotrice du clone λ 3-7

Un clone génomique de 10kb (λ 3-7, sous-cloné dans pBluescript), contenant la partie 5' flanquante du gène *reelin* souris a été précédemment isolé dans le laboratoire du Dr. T. Curran. Dans notre laboratoire, Inès Royaux a montré que ce clone contient une partie de la région 5', les exons 1 et 2 ainsi que l'intron 1 réarrangé (puisque la taille estimée de l'intron est de 80 kb et que la taille du clone est d'environ 10kb). Inès Royaux a séquencé 760 nucléotides en amont de l'ATG et déterminé, par extension d'amorces, la présence de 3 sites d'initiation de la transcription, localisés à 299, 298 et 280 nucléotides en amont du codon ATG (Royaux et al., 1997). Ces informations sont résumées à la Fig.III.1.

Dans le cadre de ce mémoire, nous avons voulu obtenir plus d'information sur la région promotrice du gène *reelin*. Dans un premier temps, nous avons déterminé la portion de région promotrice présente dans le clone λ 3-7. Par PCR « vecteur-insert », nous avons montré que ce clone détenait 1.7kb en amont de l'ATG. En effet, comme illustré à la Fig.III.2, les combinaisons de primers BSR (spécifique du vecteur pBluescript) - P2 (région promotrice *reelin*) amplifient un fragment de 1.7kb. La combinaison BSL - P2 n'amplifie aucun fragment.

Le fragment obtenu a été séquencé. La séquence obtenue ne présente aucune homologie dans les banques de données. Le clone génomique λ 3-7 étant réarrangé, nous avons voulu comparer notre séquence à celle obtenue en amplifiant le fragment à partir d'ADN génomique souris. Des primers définis dans la nouvelle séquence, nous ont permis d'amplifier un fragment de même taille. Les deux séquences sont également identiques, ce qui montre que la région obtenue n'est pas un artéfact. L'analyse de ces informations nous a permis d'obtenir une séquence de 1695bp en 5' de l'ATG, soit un « gain » de 935 nucléotides sur la séquence connue.

	CpG <i>reelin</i> souris	CpG <i>reelin</i> humain
Taille	1400bp	1200bp
% C+G	70 %	75 %
Score CpG O/E	0.98	0.93

Tableau III.1: Comparaison des îlots CpG des promoteurs *reelin* souris et humain.

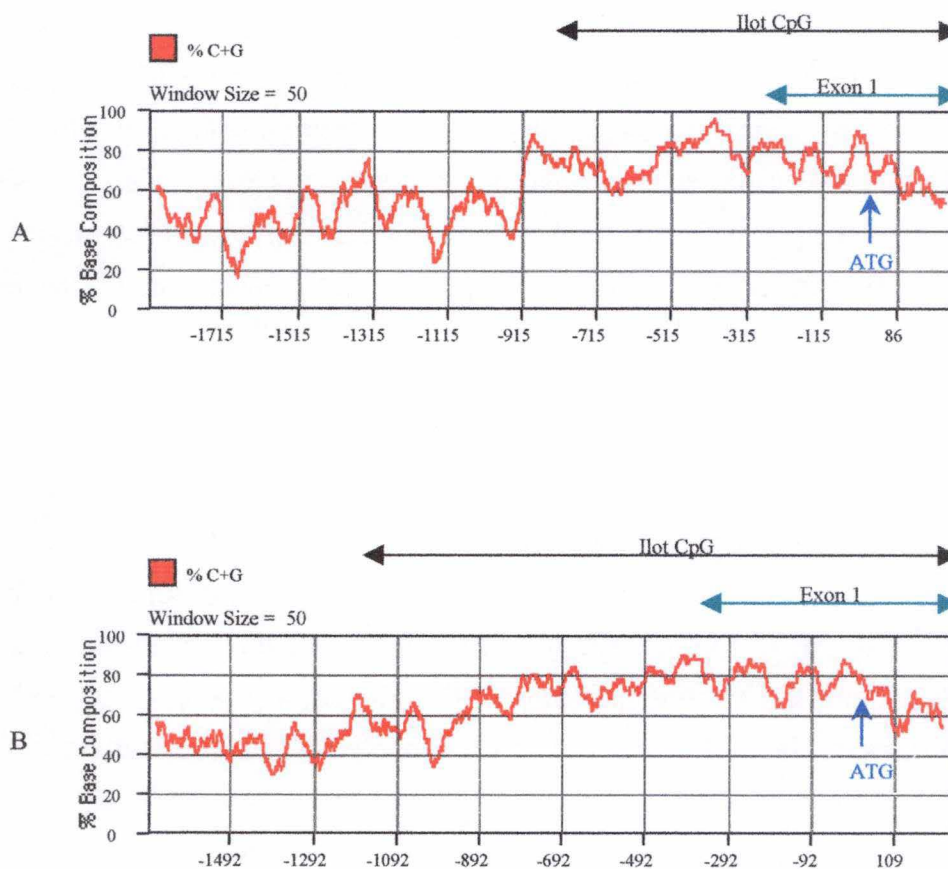


Fig.III.3: A; Pourcentage en dinucléotides CG du promoteur *reelin* humain. B; Pourcentage en dinucléotides CG présents dans le promoteur *reelin* souris. L'îlot CpG, l'exon 1 ainsi que le codon ATG sont localisés par une flèche.



Fig.III.4: Cinq régions conservées ressortent de l'alignement des séquences des promoteurs *reelin* souris et humain (programme BLAST). Les sites potentiels de liaison des facteurs de transcription (MatInspector professionnel, Genomatix, www.genomatix.de) sont encadrés ou indiqués par une flèche. Il s'agit des facteurs **XBBF** (-NNGTTRCYNNGYNACNN), **ARNT** (+GRGKATYGCGTGMSWNSCC), **IRF2** (-RSTTTCRCTTT), **IK2** (-NNNWTCCCRNNN), **AP4** (+NNCAGCTGNN), **TBPF** (-RCTTAWATAS), **PAX5** (+NCNNNRNKCANNNGNWNRKRGCSSNNN), **SP1** (-NRCCCCGCCCCCN), **WHZF** (+ANNGACGCTNN) et **AP2** (-CGCCNGSGGGMK).

-1691 gggctgtagac -1681

-1680 agtggcaggccaggcatctttagagtcagcctctgttacttggggcaagotattctgatttctcccctgtgaaatggg -1601

-1600 ggaagttttgtcttctcoggtgggtggttaggattacattaggotaatgatctacaagootgtgagcotgcatgggac -1521

-1520 tttgtctctagtgtaatattagtcgocgtgtatgaatatgtgtctgttgaactttgggaggattggaagtaatggagacc -1441

1 -1440 aacaggcagttagggtccttctctatctgtctgaaactgtgaatgtggcagaaaaaaaatcttttgcaatgaaattacac -1361

-1360 agaggctctcctttggatcttgccttctgacgccttctocccaaactcagcaagactaccattttotgacattacaaagt -1281

2 -1280 tctcgggtttctttacaagcaaggottatttttggcttgggtgtgagcotggagaaatgcattaaaacctggcagctctgtt -1201

-1200 attcccgctctctgggtgcctcggggctcctgocagggtgtctgggttaaaaaacgtgcttctgggatggtttcacttgc -1121

-1120 ccgcagtcctgctgagatagctcattaacgcgggtttcatttgtgtgocagcoactccttgcattgoccttccctccccc -1041

-1040 accccagtgacagtggtcatgcatgacatgtagaagtaataaaactcggacotaaaatttacttaaccccatctcctcccg -961

-960 aggaatcccttagcaacgcgttaatacagagtcgcggcgaagctcgcctgtcgggtccctcaccacaatcccccaggggcc -881

-880 gagaagcccgacgatctggcccccaccactccagggtggcaatcgacocgtoacccaatctcctttcagcctgggtgcc -801

-800 actgaagacccgggcccggggcagcgacgggcccggagggcctgactggaaaagggtcgcgcggggcggggcaccttcgcgg -721

3 -720 ttggggccactgacgggaggtcccccagacaagcgtgcccgtggggactggggggcgggcgcccccagggggcgagggtt -641

-640 gcagccgggtgtagacaagaacagcgcgggtcccccgcgccttcgggttctctgcgggtcctcgcagaacgggcagccgtga -561

-560 gccgggggttcgggttcccccggggcagcccggtcgggtcttggccgcttttctgttcgcgcagcgggcagcagccctgccttc -481

-480 cggggcgagcggcccggtggggcgggaggcgcgctgtgttccggggcagcccgagggtgcggggcaagcccgaggggggcg -401

4 -400 ggacggggcgccggggggcgggggcggggtgggacacgtgtggcgggcggggggagggggcgggcgccgggggtttaagaa -321

-320 ggtgcggagggggcggggcaccttcccaagcccgggcgaggggcggtcgcgtgcacacggggcgggcgggcgctcggagggcg -241

-240 gacgacgcgctctcggcgcccgcgggcccggttccccccgcgctctcgcctcggcgggcccaagtaacttcgggagcctc -161

-160 ggtctcccgctaaacttccccccggggctcgggttgcccggaacccgctcggctcagacccgcgcgcggctcggccttcccg -81

5 -80 caccgggctcctcgtgcgggtgctctccgaaagtggatgagagagcgcgggggcgcgggcgccagggagcgcgggcg -1

1 ATG GAG CGC GGC TGC TGG GCG CCG CGG GCT CTC GTC CTG GCC GTG CTG CTG CTG CTG GCG 60

1 M E R G C W A P R A L V L A V L L L L A 20

61 ACG CTG AGG GCG CGC GCG GCC ACC GGC TAC TAC CCG CGC TTC TCG CCT TTC TTT TTC CTG 120

21 T L R A R A A T G Y Y P R F S P F F F L 40

121 TGC ACC CAC CAC GGG GAG CTG GAA GGG GAT GGG GAG CAG GGC GAG GTG CTC ATT TCC CTG 180

41 C T H H G E L E G D G E Q G E V L I S L 60

181 CAC ATT GCG GGC AAC CCC ACC TAC TAC GTA CCG GGA CAG GAA TAC CAT GTT 231

Fig.III.5: Séquence du promoteur *reelin* souris. Les régions conservées (1 à 5) entre les promoteurs *reelin* souris et humain, sont indiquées en bleu.

2. Analyse des séquences et comparaison souris-homme

Ilots CpG

L'utilisation du programme MacVector 6.5, nous a permis de déceler, chez la souris, un îlot CpG de 1.4 kb s'étendant à la fois sur la région promotrice et l'exon 1 (Fig.III.3a) : plus de 70 % de GC, présence de nombreux sites de restriction associés aux îlots CpG (*MluI*, *BssHII*, ...). Ceci est en accord avec la carte génomique présentée à la Fig.I.6. Afin d'identifier des éléments importants pour la régulation de la transcription, nous avons décidé de comparer la séquence souris à celle du promoteur *reelin* humain. Aucune séquence d'autres espèces n'est disponible actuellement. La séquence complète (162912nt) d'un clone BAC humain contenant 160kb de région 5', l'exon 1 et 2.7kb de l'intron 1 du gène *reelin*, est accessible dans GenBank (AC002067). Chez l'homme, un îlot CpG de 1.2 kb couvre la région promotrice et s'étend jusqu'à une centaine de nucléotides à l'intérieur de l'intron 1 (Tableau III.1 et Fig.III.3b). Les promoteurs *reelin* souris et humain sont dépourvus de boîtes TATA et CCAAT. Nous n'avons pas pu identifier de sites consensus Inr, DPE ni BRE. En résumé, aussi bien chez l'homme que chez la souris, le promoteur proximal de *reelin* ne contient pas de boîte TATA et est associé à un îlot CpG.

Sites potentiels de fixation de facteurs transcriptionnels

Afin d'identifier des sites de liaison pour des facteurs de transcription qui joueraient un rôle dans la régulation de l'expression de *reelin*, nous avons comparé les deux séquences. Nous pensons, en effet, que des sites importants doivent être conservés au cours de l'évolution. L'alignement des séquences des promoteurs humain et souris, est présenté à la Fig.III.4. Six régions sont relativement bien conservées. La région couvrant le site d'initiation de la traduction est extrêmement bien conservée (82 %). Nous avons ensuite voulu déterminer si les cinq autres régions conservées contenaient des sites potentiels de fixation pour des facteurs de transcriptions impliqués dans le développement. Les régions numérotées de 1 à 5, reprises à la Fig.III.5, ont été analysées au moyen du programme MatInspector professionnel chez Genomatix (www.genomatix.de). Les résultats sont présentés dans le tableau III.2.

	Name of family/matrix	Further information	Strand	Core sim.	Matrix sim.	Sequence
1	V\$FKHD/FREAC4.01	Fork head related activator-4	(-)	0,750	0,781	cctctgtatATCActtc
2	V\$XBBF/RFX1 01	X-box binding protein RFX1	(+)	1,000	0,906	ttaaacctgGCAAccc
	V\$NFKB/NFKAPPAB65 01	NF-kappaB (p65)	(-)	0,782	0,862	cggggtTGCC
	V\$ETSF/CETS1P54 01	c-Ets-1(p54)	(+)	1,000	0,958	tcCGGAtgcc
	V\$CREB/TAXCREB 02	Tax/CREB complex	(-)	1,000	0,710	ggagcccTGCgcat
	V\$EBOX/ARNT 01	AhR nuclear translocator homodimers	(+)	1,000	0,884	ttaaaaaCGTGcttat
	V\$CLOX/CLOX 01	Clox	(+)	0,804	0,913	cttATGgattatttc
	V\$IRFF/IRF2 01	Interferon regulatory factor 2	(-)	1,000	0,836	gggaagtGAAAta
3	V\$AP2F/AP2 Q6	Activator protein 2	(-)	0,976	0,866	cgCCCCtcgggg
	V\$IK2 01	Ikaros 2	(+)	1,000	0,902	gcgcGGGAcgca
	V\$CETS1P54 01	c-Ets-1(p54)	(+)	0,833	0,833	gcGGGAcgca
	V\$AP4 Q6	Activator protein 4	(+)	1,000	0,887	cgCAGCcgge
	V\$AP4 Q5	Activator protein 4	(+)	1,000	0,911	cgCAGCcgge
	V\$HFH3 01		(-)	0,838	0,889	gttTCTTtgtctg
4	V\$TBPF/MTATA B	Muscle TATA box	(-)	1,000	0,857	cttctTAAAgccccggg
	V\$PAX5/PAX5 02	B-cell-specific activating protein	(+)	1,000	0,747	ttaagaaggtgtggAGCGgggc gggcgc
	V\$PAX1/PAX1 B	Pax 1 binding sites	(-)	0,750	0,649	CCGCcccgctccacact
	V\$SPIF/SP1 Q6	Stimulating protein 1	(+)	1,000	0,926	gcggGGCGggcgc
	V\$EGFR/WT1 B	Milms Tumor suppressor	(+)	0,796	0,871	ggcCGGgctttt
	V\$IKRS/IK2 01	Ikaros 2	(-)	1,000	0,971	gcctGGGAaage
	V\$AP2F/AP2 Q6	Activator protein 2	(-)	0,976	0,905	cgCCCCtcggcc
	V\$EGFR/NGFIC 01	Nerve growth factor- induced protein C	(+)	0,787	0,820	agGCGGcgggcg
	V\$WHZF/WHN B	Winged helix	(+)	1,000	0,936	gacgACGCgct
	V\$AP2F/AP2 Q6	Activator protein 2	(+)	1,000	0,908	cgCCCgcagccc
	V\$AP2F/AP2 Q6	Activator protein 2	(+)	1,000	0,892	ctCCCgcggccc
	V\$XBBF/RFX1 01	X-box binding protein RFX1	(+)	0,945	0,891	gcggcccaaaGTAActt
5	V\$IK1 01	Ikaros 1	(-)	1,000	0,860	tgcgGGGAaggcg
	V\$IK2 01	Ikaros 2	(-)	1,000	0,914	tgcgGGGAaggc
	V\$MZF1 01	Myeloid zinc finger 1	(-)	1,000	0,975	tgCGGGa
	V\$AHRARNT 01	Ah receptor nuclear translocator	(-)	1,000	0,913	gaggagccgCGTGcgg
	V\$NMYC 01		(-)	1,000	0,893	agccgCGTGcgg
	V\$USF Q6	Upstream stimulatory factor	(+)	1,000	0,903	cgCACGcgge
	V\$USF C	Upstream stimulatory factor	(-)	0,813	0,873	cCGCGTgc
	V\$LYF1 01	LyF-1	(-)	0,763	0,863	tttCGGAagg
	V\$NKX2.5 01		(+)	1,000	0,884	gaAAGTg
	V\$TCF11 01	T cell factor	(-)	0,798	0,864	CTCAtcacatttc

Tableau III.2: Liste des sites potentiels de fixation des facteurs de transcription relatifs aux séquences homologues.

Localisation sur la séquence du promoteur reelin humain; 1: -1538 à -1499. 2: -1367 à -1261. 3: -730 à -771. 4: -344 à -161. 5: -94 à 40

De notre analyse, nous avons retenu les sites conservés chez l'homme et la souris et qui présentaient une homologie de 100 % avec le site minimal de liaison du facteur (Core sim.=1). Ces sites sont les suivants : X-box binding protein RFX1, AhR nuclear translocator homodimers, Interferon regulatory factor 2, Ikaros2, Activator protein 4, Muscle TATA box, B-cell-specific activating protein, Stimulating protein 1, Winged helix et Activator protein 2.

3. Criblage d'une banque d'ADN génomique de souris

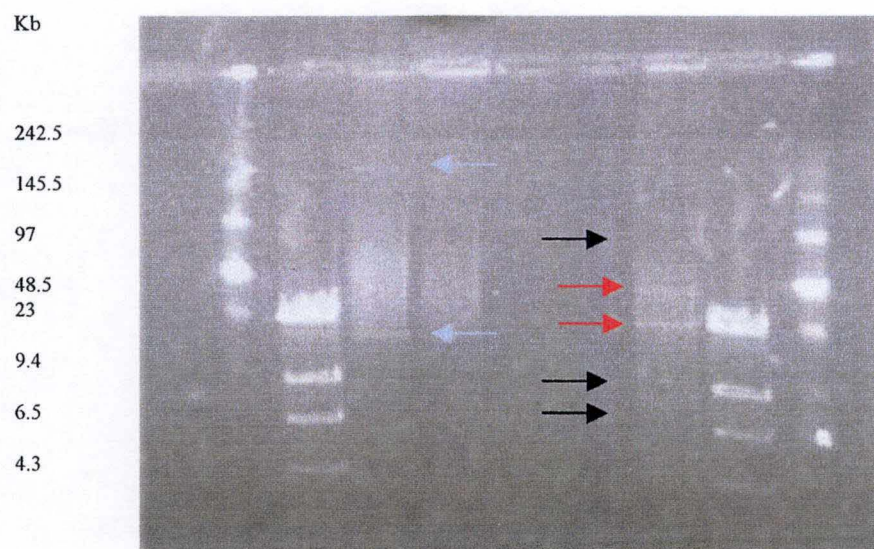
Il nous semblait intéressant de disposer de plus de séquence en 5', et nous avons donc décidé de cribler une banque d'ADN génomique. Les banques génomiques en BAC ou PAC représentent actuellement le meilleur système pour isoler de grandes portions de génome. Les BACs (Bacterial Artificial Chromosome) ont pour base l'épisome F de 7 kb d'*E.coli*, et les PACs (P-1 derived Artificial Chromosome) utilisent un vecteur dérivé du bactériophage P-1. C'est avec ces derniers que nous avons travaillé. Le vecteur P-1 est présent à l'état de copie unique dans *E.coli*, ce qui permet d'éviter les recombinaisons. Les PACs peuvent facilement être amplifiés en vue de préparer de l'ADN. La taille des inserts varie entre 100 et 300 kbp.

Des banques PAC souris ont été construites par les Dr de Jong et Osoegawa, et sont disponibles pour un prix modéré, au RZPD (Berlin), sous forme de pools d'ADN pour criblage par PCR.

Nous avons utilisé la banque N°711, produite à partir d'ADN de rate de souris. Cette banque couvre 13 fois la taille du génome et possède des inserts d'une taille moyenne de 146kb. La banque est organisée en super-pools et pools pour un criblage facile en PCR. La banque est criblée en réalisant 84 (super-pools)+ 48 (pools) PCR (voir Fig.II.5). Cette banque a été criblée au moyen d'amorces spécifiques de la région promotrice souris (P1 et P2). Nous avons obtenu le clone positif 506C10.

i

λ λHindIII PAC (NotI) PAC (NotI-MluI) λHindIII λ



ii

Fragments reconnus par la sonde P2-P3

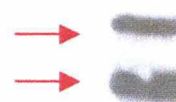


Fig.III.6:i Electrophorèse sur gel d'agarose dans un champ électrique pulsé. PAC (NotI), restriction de l'ADN de PAC par NotI qui permet la libération de l'insert. Les flèches mauves indiquent la position de l'insert (160kb) et du vecteur (20kb). PAC (NotI-MluI), restriction par les deux enzymes précitées. Cinq fragments sont détectés (flèches noires et rouges). Un Southern blot a été réalisé à partir de cette piste.

ii: hybridation avec une sonde couvrant une partie de la région promotrice. Les deux bandes, de 20 et 50kb, correspondent à celles pointées par les flèches rouges dans la figure i.

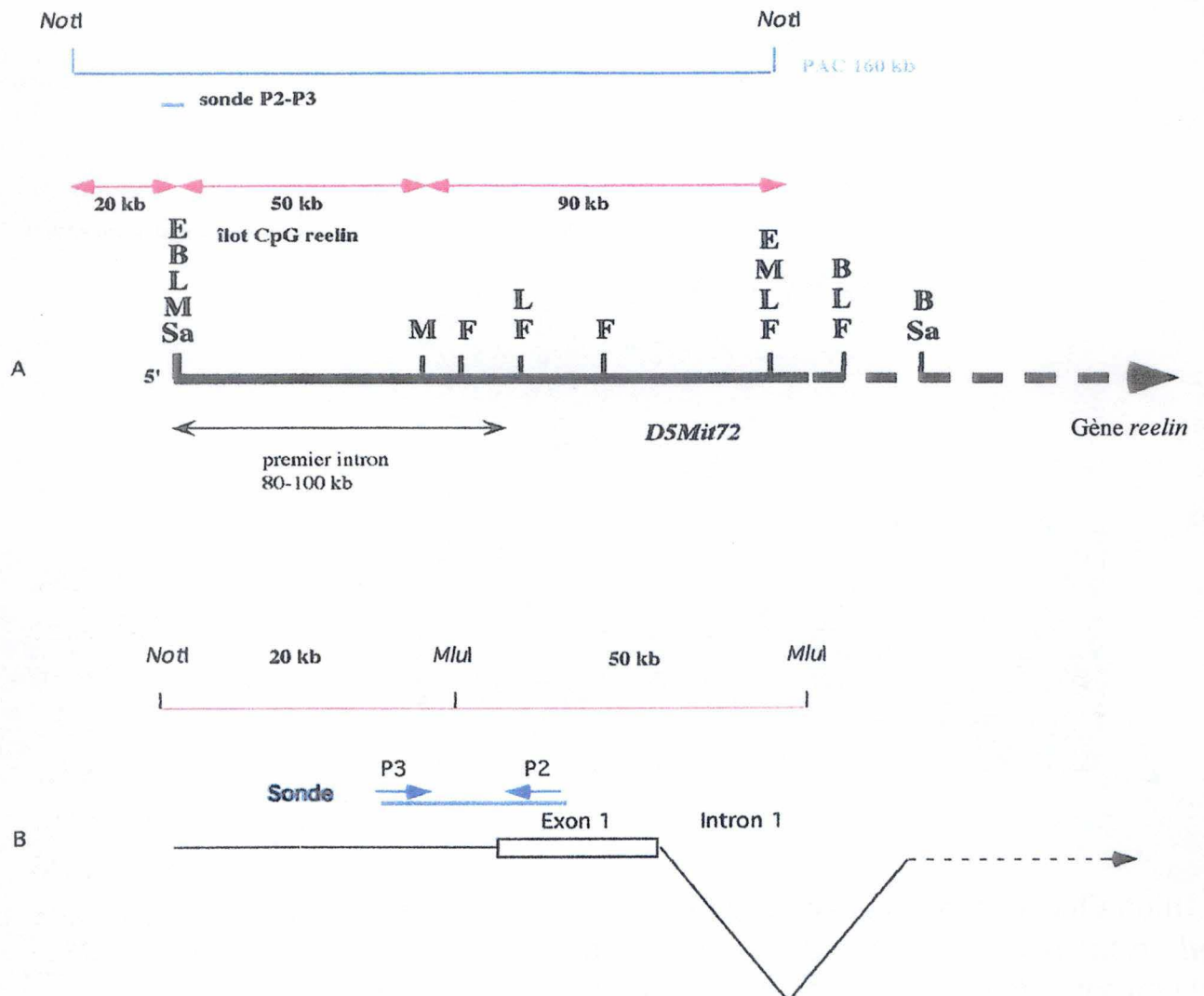


Fig.III.7: Caractérisation du clone obtenu par criblage d'une banque PAC d'ADN génomique de souris.

A; Les régions contenant un groupe de sites pour des enzymes de restriction typiques des îlots CpG sont indiquées sur le gène *reelin* par les lettres suivantes; E=*EagI*, B=*BssHII*, L=*SalI*, M=*MluI*, Sa=*SacII* et F=*SfiI*. Après double digestion *NotI*-*MluI*, un Southern blot réalisé avec une sonde P2-P3 (couvrant 1.7kb de la région promotrice), a permis de révéler deux bandes (20 et 50kb).

B; Aggrandissement de la région promotrice, de l'exon 1 et de l'intron 1. Ce dessin permet de localiser la position de la sonde, ainsi que les sites *NotI* et *MluI*.

Caractérisation du clone positif

a. Taille de l'insert

Le site de clonage du vecteur pPAC4 est bordé par deux sites *NotI*, ce qui permet de libérer l'insert en un seul fragment (ce site est très rare dans le génome et nous savons qu'il est absent de la région promotrice de *reelin*). Après digestion du PAC par *NotI* et électrophorèse en champ pulsé, nous obtenons deux fragments (Fig.III.6i) : un fragment de 20kb correspondant au vecteur et un fragment de 160kb. Cette information a été vérifiée par Southern blot : après transfert sur membrane de nylon et hybridation avec une sonde de la région promotrice, nous reconnaissons bien le fragment de 160kb.

b. Contenu du PAC

Nous nous sommes basé sur la carte génomique du gène *reelin* (Fig.III.7, Bar et al., 1995), et la séquence de la région 5' (Fig.III.5) : un site *MluI* étant présent en position -945, nous avons digéré le PAC par *NotI* (dans le vecteur) et *MluI* (dans le promoteur). Après électrophorèse en champ pulsé, nous avons identifié cinq fragments de 7 kb, 10 kb, 23 kb, 48 kb et 100 kb (le vecteur contient plusieurs sites *MluI* et est donc digéré en plusieurs petits fragments). Afin de déterminer quelle bande correspond à la région promotrice, le gel a été transféré sur membrane de nylon et hybridé avec la sonde P2-P3 couvrant cette région et contenant le site *MluI*. Deux bandes de 23 et 50 kb ont été reconnues. En se basant sur les données de la carte génomique du gène *reelin* (et la position des sites *MluI* sur cette carte) nous pensons que le fragment de 20 kb est un fragment *NotI-MluI* contenant la région promotrice, et que le fragment de 50 kb est celui situé entre le site *MluI* de la région promotrice et celui situé dans le premier intron (Fig.III.7). Puisque la distance entre le vecteur et le promoteur *reelin* est de 20 kb, et que la taille du PAC est de 160 kb, notre clone contient encore 140 kb de *reelin*. Nous pensons donc que outre la région promotrice, l'exon 1 et l'intron 1, ce PAC contient plusieurs autres exons. Nous avons réalisé des PCR au moyen des amorces spécifiques des exons 2, 3 et 4. Ces PCR ont été positives. Au moment de la rédaction de ce travail, nous n'avons pas

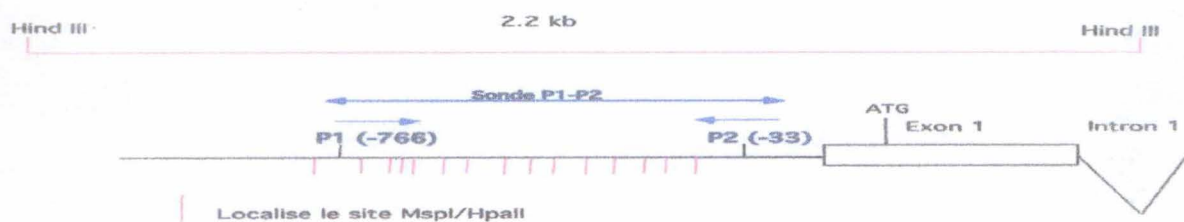


Fig.III.8.a: La région promotrice et l'exon 1 sont contenus dans un fragment *Hind*III de 2.2 kb. 15 sites pour l'enzyme *Msp*I (clivant le site c/cgg indépendamment de son état de méthylation) et son isoschisoisomère *Hpa*II (clivant le site c/cgg non méthylé) sont localisés par une barre verte. Après hybridation avec la sonde P1-P2, si les sites sont méthylés un fragment de 2.2kb est obtenu par la digestion *Hind*III/*Hpa*II, tandis que 15 fragments allant de 8 à 150 bases sont engendrés par la digestion *Hind*III/*Msp*I. S'il n'y a pas de méthylation, 15 fragments sont engendrés par les deux digestions.

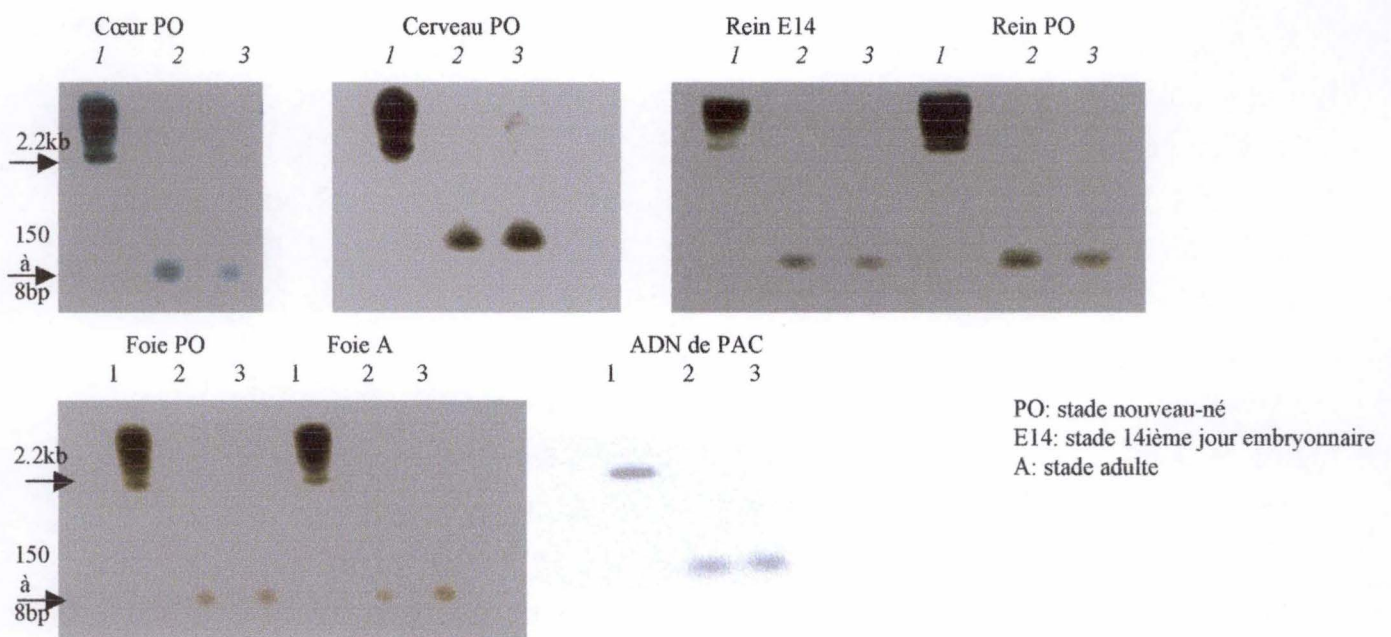


Fig.III.8.b: Par Southern blot, nous avons montré que la région promotrice et l'exon 1 sont contenus dans un fragment *Hind*III de 2.2 kb. Piste 1; digestion *Hind*III, piste 2; double digestion *Msp*I/*Hind*III, piste 3; double digestion *Hpa*II/*Hind*III. Le pattern obtenu pour ces digestions est identique à partir de l'ADN des 4 organes étudiés aux différents stades illustrés. Comme contrôle positif, l'ADN du PAC obtenu lors du criblage, a été digéré par les mêmes couples d'enzymes. A la piste 1, nous observons une bande de 2kb; les autres pistes contiennent un « smear » de bandes allant de 8 à 150 bases. Le site c/cgg n'est donc pas méthylé, du moins dans l'ADN des organes étudiés.

déterminé avec exactitude si d'autres exons localisés plus en 3' sont présents dans le PAC.

En résumé, nous avons isolé un clone PAC qui contient 20 kb de région promotrice, ainsi que l'exon 1, l'intron 1 d'environ 80 kb et plusieurs exons localisés plus en 3'. À ce stade, nous ne pensons pas que ce clone est réarrangé.

4. Etude de la régulation de l'expression par la méthylation

L'intervention de la méthylation dans la régulation de l'expression reste controversée. Cependant, d'aucuns proposent que, dans le cas des gènes à expression spécifiques, un îlot CpG couvrant la région promotrice ne serait pas méthylé dans les tissus où le gène est exprimé mais l'est dans les tissus où le gène est éteint. *Reelin* étant un gène à l'expression spécifique, nous avons étudié son état de méthylation dans divers organes à différents stades du développement.

Nous avons choisi la méthode consistant à digérer l'ADN génomique par des enzymes sensibles à la méthylation. L'ADN de cerveau, de cœur, de rein et de foie a été extrait au 14^{ième} jour embryonnaire (E14) ainsi qu'au stade adulte (A) et nouveau-né (P0).

Dans une première étape, par Southern blot, nous avons montré que la région promotrice et l'exon 1 étaient contenu dans un fragment *Hind*III de 2.2 kb. L'îlot CpG, contenu dans ce fragment de 2.2 kb, contient 15 sites pour l'enzyme *Msp*I et son isoschisoisomère *Hpa*II. Ce dernier clive le site c/cgg uniquement si celui-ci est non méthylé et *Msp*I clive le site c/cgg, indépendamment de son état de méthylation. 5 µg d'ADN, à chaque stade, sont digérés par *Hind*III seul, par le couple *Hpa*II/*Hind*III et par *Msp*I/*Hind*III. Si les sites sont méthylés, nous pourrions observer un fragment de 2.2 kb pour la digestion *Hpa*II/*Hind*III, tandis que nous obtiendrons de nombreux petits fragments (de 8 bases à 150 bases) pour la seconde digestion après hybridation avec une sonde P1-P2 (Fig.III.8a). Comme contrôle positif, l'ADN du PAC obtenu lors du criblage, a été digéré par les mêmes couples d'enzymes.

Comme illustré à la Fig.III.8b, le pattern obtenu pour la digestion *Hind*III-*Msp*I et *Hind*III-*Hpa*II est identique. Ce résultat montre : premièrement, qu'il s'agit bien d'un vrai îlot CpG puisque cette région n'est pas méthylée. Et

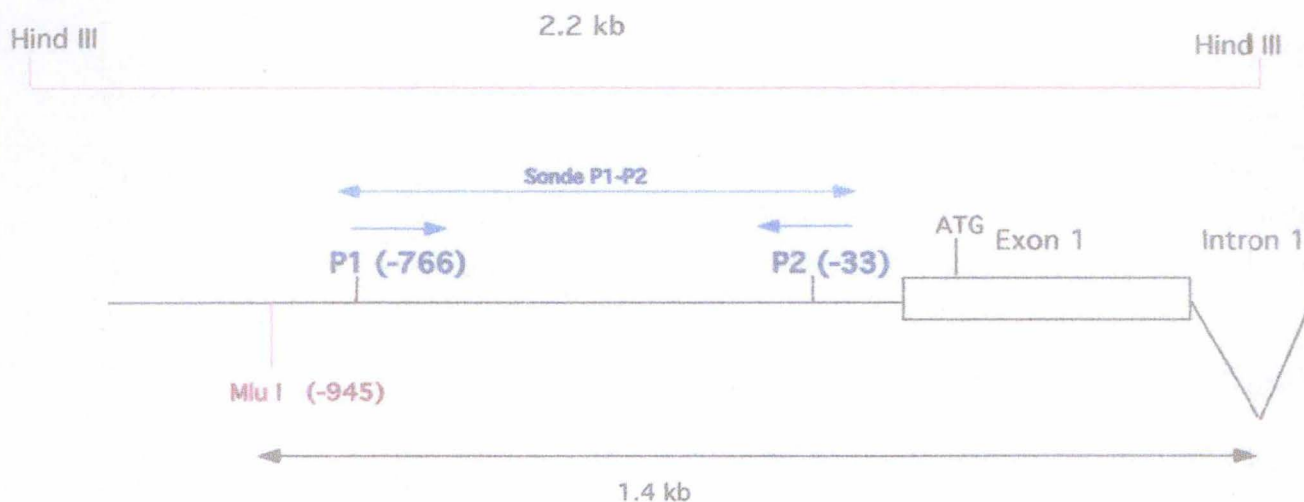
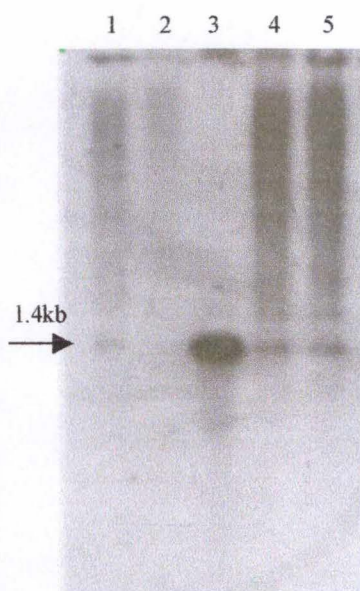


Fig.III.9a: *Mlu*I est un enzyme typique des îlots CpG. Il clive le site a/cgcgt uniquement si celui-ci n'est pas méthylé. Un site *Mlu*I est localisé à -945bp. De plus, une digestion *Hind*III engendre un fragment de 2.2kb contenant la partie promotrice, l'exon 1 et un fragment de l'intron 1. Si le site a/cgcgt n'est pas méthylé, nous observons un fragment de 1.4kb après digestion *Hind*III/*Mlu*I, après hybridation avec une sonde P1-P2. Dans le cas contraire, cette sonde reconnaît un fragment de 2.2kb.



- 1 Cœur A (=adulte), double digestion *Hind*III-*Mlu*I.
- 2 Cœur E14 (=14ième jour embryonnaire), double digestion *Hind*III-*Mlu*I.
- 3 ADN de PAC, double digestion *Hind*III-*Mlu*I.
- 4 Cerveau A, double digestion *Hind*III-*Mlu*I.
- 5 Cerveau E14, double digestion *Hind*III-*Mlu*I.

Fig.III.9b: Southern blot réalisé après digestion *Hind*III-*Mlu*I. Le pattern est identique pour l'ADN du cœur et du cerveau, prélevés aux stades E14 et Adulte. Comme contrôle positif, l'ADN du PAC obtenu lors du criblage, a été digéré par le même couple d'enzymes. Nous pouvons observer sur la photo une bande de 1.4kb. Le site *Mlu*I (a/cgcgt) n'est donc pas méthylé.

deuxièmement que la méthylation ne joue aucun rôle dans l'expression tissu spécifique de *reelin*, au moins dans les tissus choisis et aux stades étudiés.

Nous avons également vérifié l'état de méthylation d'autres sites : l'ADN a été digéré par les couples *MluI-HindIII*, *BssHII-HindIII*, *EagI-HindIII*, *SmaI-HindIII* et *SacII-HindIII*. La figure 9 illustre les résultats obtenus pour le couple *MluI-HindIII*. Si le site étudié est méthylé, nous devons obtenir un fragment de 2.2 kb. Dans le cas contraire, nous obtiendrons un fragment d'environ 1.4 kb (Fig.III.9a). Après analyse en Southern blot, nous obtenons bien un fragment de 1.4 kb dans les différents tissus testés.

Bien qu'instructive, cette approche limite notre champ d'investigation à certains sites de la région étudiée, de plus les sites hémi-méthylés ne sont pas détectés. Il serait intéressant de réaliser l'expérience en utilisant la technique basée sur la capacité du bisulfite de sodium de désaminer efficacement les résidus cytosines (non méthylés) en uraciles. L'ADN est ensuite amplifié par PCR. Les résidus uraciles et thymines sont amplifiés en thymines, et les résidus 5-MeC en cytosines. Après clonage dans un vecteur approprié, l'ADN est séquencé (Clark et al.,1994).

5. Etude in vitro du promoteur *reelin* souris

Notre étude a été réalisée avec une lignée cellulaire nommée P19, dérivée d'un teratocarcinome de souris C3H/He, qui sont un bon modèle pour analyser la régulation de la différenciation neuronale. Le traitement à l'acide rétinoïque d'un agrégat de P19, entraîne leur différenciation en neurones (Bain et al., 1994). Pour une question de temps, nous n'avons pu exploiter cette propriété. Nous avons montré par RT-PCR que les cellules P19 expriment *reelin*.

Nous avons construit différents plasmides correspondant à des délétions progressives de la région promotrice *reelin* (Fig.III.10). Une PCR avec le couple d'amorce BSR-P2 a permis d'amplifier la région promotrice du clone λ 3-7 (fragment de 1.7kb) qui a été sous-cloné dans le vecteur pCR2.1 (kit TA Cloning). Ce fragment de 1.7kb ainsi que des sous-régions ont été clonés dans le vecteur pGL3 basic. Ce vecteur possède un ADNc de la luciférase *firefly* comme gène rapporteur mais ne possède pas de promoteur minimal. Lorsque le

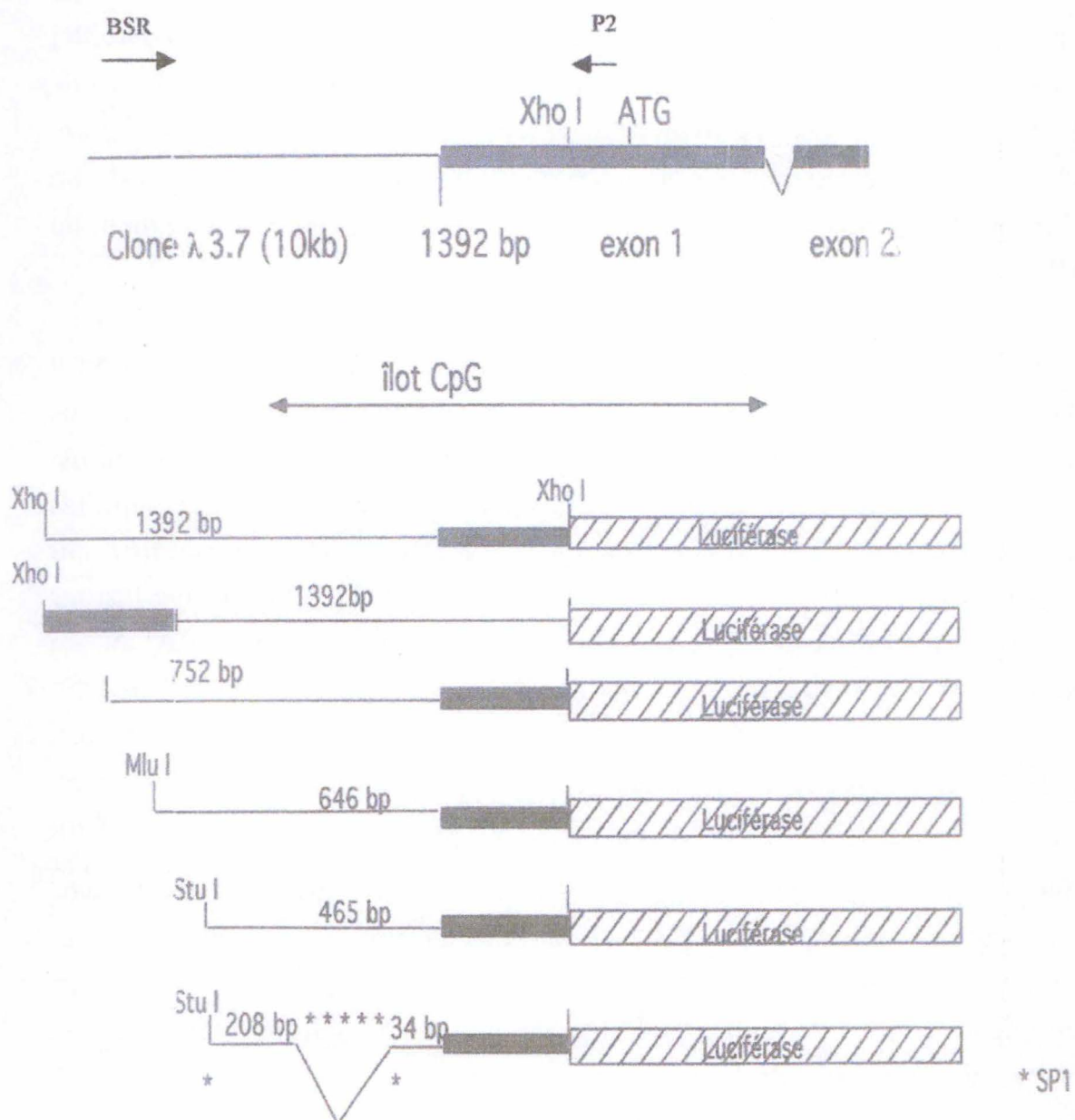


Fig.III.10: Constructions génétiques du promoteur *reelin* souris.

fragment cloné en amont du gène rapporteur contient un promoteur, le gène *luciférase* est exprimé. Cinq constructions ont été réalisées. Trois d'entre elles ont été obtenues par digestion ; *XhoI-XhoI*, *MluI-XhoI* et *StuI-XhoI*. Il s'agit, respectivement, des constructions désignées : Rls 1392bp -, Rls 646bp et Rls 465bp. Elles contiennent 7 sites de liaisons potentiels au facteur Sp1. Par restriction avec *SmaI*, nous en avons enlevé 5. La construction se nomme Rls délétion. Un dernier fragment, nommé Rls 752bp, a été obtenu de manière fortuite au départ du clone Rls 1392bp -> (Fig.III.10). Ces constructions ont été vérifiées par restrictions et séquençage.

Dans une seconde étape, nous avons transfecté les clones (pGL3 + construction) dans la lignée cellulaire P19 en utilisant la Lipofectamine 2000. Trois contrôles ont été transfectés. Le vecteur pGL3 basic seul permet de déterminer le blanc. Rls1392 <- a été cloné en sens inverse de la transcription ; il ne peut y avoir expression de la luciférase. Enfin, le contrôle CNT (cellules non transfectées) nous permet de déterminer le bruit de fond. Pour la lecture de l'activité luciférase, nous avons utilisé le kit Dual Luciferase Reporter Assay System, (Promega) dont le principe est le suivant : une cotransfection d'un plasmide test « *Firefly* » et d'un plasmide contrôle « *Renilla* », dont la régulation ne varie pas, permet à la fois de quantifier et de normaliser les expériences. Brièvement, les plasmides possèdent un gène luciférase provenant de deux organismes différents. Les deux luciférases ont des structures différentes et interagissent avec des substrats différents pour leur réaction en bioluminescence. Nous mesurons donc deux activités luciférases dans chaque échantillon : la *firefly* (pour le test) et la *Renilla* luciférase (sous contrôle d'un promoteur TK). L'activité de la *Renilla*, sous contrôle TK, varie selon l'efficacité des transfection et sert donc de normalisation.

Nous avons réalisé 4 séries de transfections dont les résultats sont les suivants ;

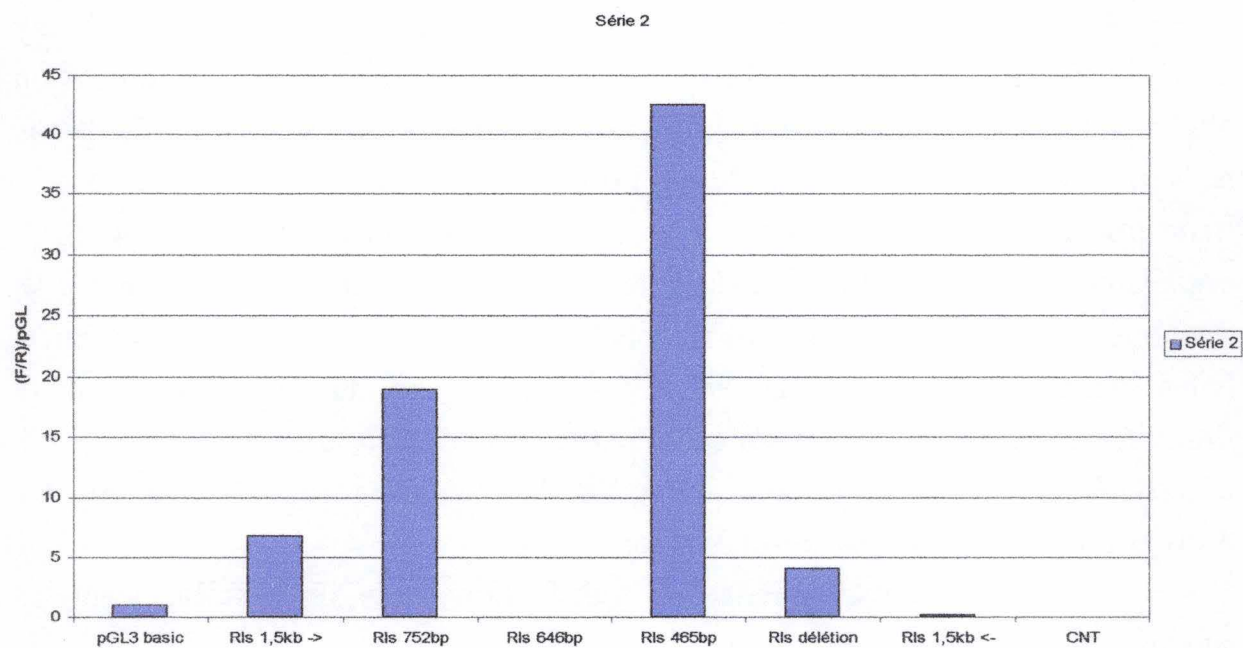
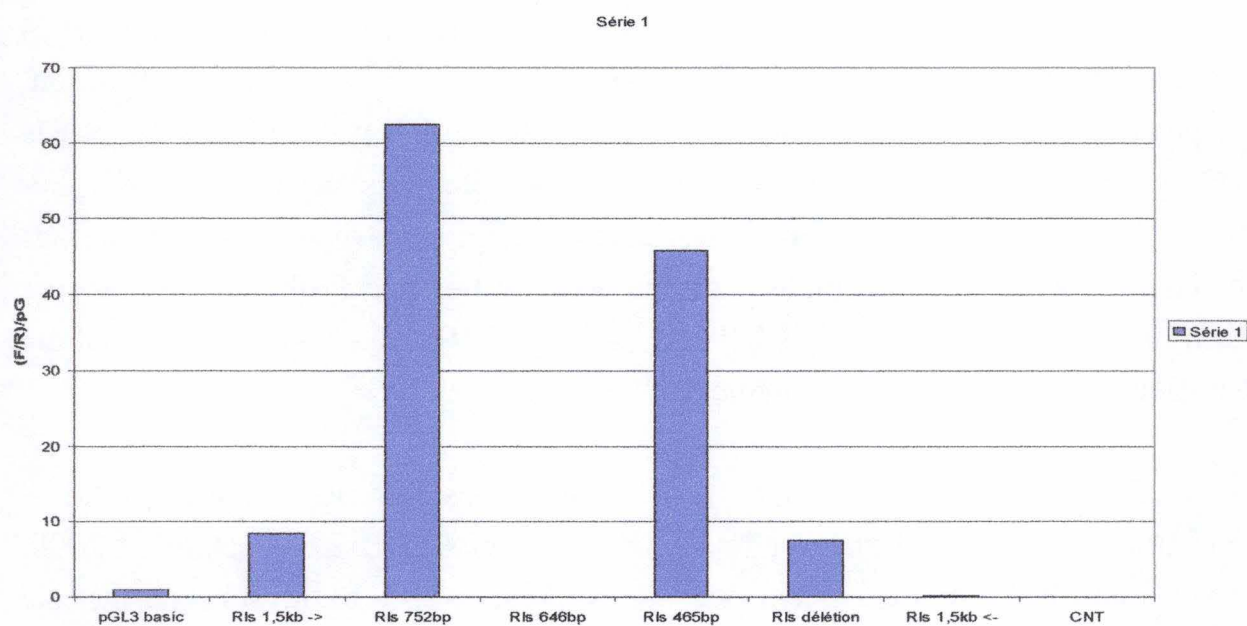


Fig.III.11: Activité luciférase induite par diverses constructions du promoteur *reelin* de souris, par rapport au vecteur pGL3 seul et à des cellules non transfectées (CNT).

Série 1

Construction	µl testé	Firefly	Renilla	F/R	(F/R) / pGL3
pGL3 basic	10	40	227	0.43243243	1
Rls1392bp ->	10	260	223	3.6	8.325
Rls 752bp	10	170	159	27	62.4375
Rls 646bp	10	93	75		
Rls 465bp	10	1057	206	19.7924528	45.7700472
Rls délétion	10	165	202	3.20408163	7.40943878
Rls1392bp <-	10	15	282	0.05426357	0.1254845
CNT	10	8	153		

L'activité de la luciférase est exprimée en RLU (relative luciferase units).

(F/R)/pGL3 nous permet de déterminer l'intensité de l'expression de la luciférase par rapport au vecteur seul. Cela indique la « force » du promoteur. F/R représente le ratio de l'intensité lumineuse produite par la luciférase *Firefly* à celle produite par la luciférase *Renilla*. Lors de cette opération, nous avons soustrait à toutes les valeurs, celle du bruit de fond.

Série 2

Construction	µl testé	Firefly	Renilla	F/R	(F/R) / pGL3
pGL3 basic	10	31	176	0.30666667	1
Rls1396bp ->	10	332	257	2.07692308	6.77257525
Rls 752bp	10	789	235	5.82835821	19.0055159
Rls 646bp	10	197	83		
Rls 465bp	10	1419	209	13.0648148	42.602657
Rls délétion	10	215	269	1.23214286	4.01785714
Rls1396bp <-	10	11	154	0.05660377	0.18457752
CNT	10	8	101		

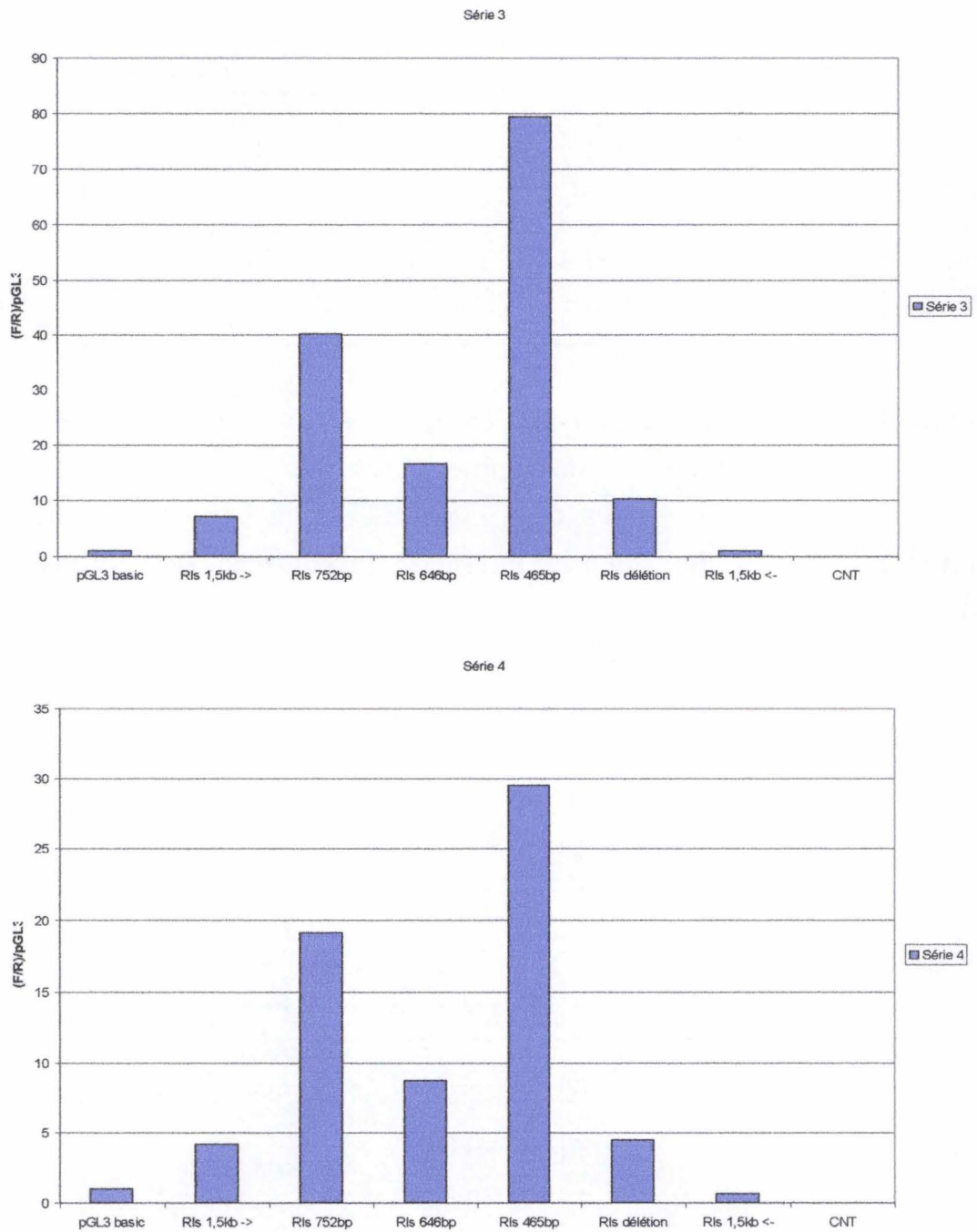


Fig.III.11: Induction de l'activité luciférase par diverses constructions du promoteur *reelin* de souris (RIs). pGL3 basic: vecteur sans insert. CNT: contrôle non transfecté.

Série 3

Construction	µl testé	Firefly	Renilla	F/R	(F/R) / pGL3
pGL3 basic	10	81	448	0.2084507	1
Rls1396bp ->	10	557	459	1.50273224	7.20905332
Rls 752bp	10	1207	236	8.39160839	40.2570403
Rls 646bp	10	922	356	3.47908745	16.6902168
Rls 465bp	10	2010	214	16.553719	79.4131115
Rls délétion	10	443	297	2.1372549	10.2530472
Rls1396bp <-	10	74	411	0.21069182	1.01075132
CNT	10	7	93		

Série 4

Construction	µl testé	Firefly	Renilla	F/R	(F/R) / pGL3
pGL3 basic	10	21	203	0.11184211	1
Rls1396bp ->	10	269	613	0.47153025	4.21603517
Rls 752bp	10	654	355	2.13815789	19.1176471
Rls 646bp	10	188	239	0.9787234	8.75093867
Rls 465bp	10	724	269	3.30275229	29.5304911
Rls délétion	10	132	306	0.50196078	4.48811995
Rls1396bp <-	10	15	204	0.07189542	0.64282968
CNT	10	4	51		

Les constructions du promoteur *reelin* souris se comportent dans les quatre séries de façon homogène. Nous n'avons pas tenu compte des valeurs obtenues dans les séries 1 et 2, pour la construction nommée Rls 646bp. En effet, la valeur du bruit de fond (CNT) est plus élevée que celle du test. Malgré cela, on constate une augmentation de l'expression du gène rapporteur dans le cas des constructions Rls 752bp et Rls 465bp, et une diminution de l'expression avec la construction Rls délétion (242bp), et ce dans les quatre séries (Fig.III.11).

Bien que préliminaires, ces résultats suggèrent la présence d'un site pour un répresseur dans la construction 1.5kb, en amont du nucléotide -752. Il en est de même pour la construction 646bp, où le site pour un élément de régulation négative se situerait entre le nucléotide -646 et -465. Enfin, les facteurs Sp1 semblent être d'une grande importance pour la régulation de la transcription du gène *reelin*. En effet, la délétion de 5 sites sur les 7 présents diminue très fortement l'expression du gène rapporteur. Les résultats sont à considérer comme préliminaires car, d'une part, les cellules P19 sont difficiles à transfecter d'où une efficacité de transfection faible et d'autre part, pour une même construction, la différence entre les valeurs de luciférase obtenues est importante. Cependant, dans les 4 séries de transfections, les résultats vont dans le même sens. Sans résultats expérimentaux supplémentaires, il est difficile de dégager l'une ou l'autre hypothèse concernant un répresseur potentiel.

*Conclusions,
discussion et
perspectives,*

IV. Conclusions, discussion et perspectives

Nous avons déterminé la séquence de 1.7 kb de la région promotrice du gène *reelin* de souris. Cette analyse nous permet de confirmer que ce promoteur est dépourvu de boîte TATA mais fait partie d'une région qui possède toutes les caractéristiques d'un îlot CpG, avec un contenu en G+C de 70 % dont 98 % de doublets CpG. Cet îlot couvre la région promotrice et environ 500 nucléotides de l'exon 1. *Reelin* possède plusieurs sites d'initiation de la transcription, ce qui est aussi une caractéristique des gènes associés aux îlots CpG et ne possédant pas de boîte TATA. Enfin, cette région contient de nombreux sites pour des enzymes de restriction coupant rarement (15 sites pour l'enzyme *HpaII*), particularité typique des îlots CpG.

Ce type de promoteur est classiquement rencontré dans les gènes de maintenance, mais on en découvre de plus en plus dans des gènes à expression spécifique comme *reelin*, par exemple les gènes codant pour la β 4-intégrine (Takaoka et al., 1998), l'aldolase C spécifique du cerveau (Buono et al., 1993), ou encore l' α -globine (Bird, 1987).

La présence d'un îlot CpG dans la région promotrice pose la question du rôle de la méthylation dans l'expression de *reelin*. La régulation de l'expression par la méthylation est un sujet controversé. Une hypothèse classique est que la méthylation permet de contrôler l'expression d'un gène : promoteur méthylé dans les tissus qui n'expriment pas le gène et promoteur déméthylé dans les tissus qui l'expriment. *Reelin* étant un gène à expression tissu spécifique, nous avons examiné son pattern de méthylation et nos résultats ne cadrent pas du tout avec cette vision classique. En effet, les sites étudiés (c/cgg, clivé par l'enzyme *HpaII* et par son isoschisoisomère *MspI*, ainsi que le site a/cgcgt, clivé par l'enzyme *MluI*) ne montrent aucune variation de méthylation, que ce soit au niveau du cœur, du foie, du rein et du cerveau prélevés à différents stades de développement, 14^{ème} jour embryonnaire, nouveau-né et adulte. La méthylation ne jouerait donc aucun rôle dans l'expression du gène *reelin* du moins dans les tissus aux stades étudiés. Bien qu'instructive, cette approche est limitée à

certains sites de la région étudiée, et les sites hémi-méthylés ne sont pas détectés. Il serait donc nécessaire de compléter notre analyse par la technique de séquençage après conversion des cytosines en uraciles par le sodium bisulfite. Toutefois, il semble bien que l'hypothèse classique du rôle de la méthylation dans la régulation de l'expression soit actuellement revue et que la méthylation joue surtout un rôle dans l'extinction définitive d'un gène au stade précoce du développement davantage que dans la régulation réversible de son expression dans certaines cellules.

Etant donné l'expression très particulière de *reelin* dans le cerveau, en particulier l'amplification énorme dans les cellules de Cajal-Retzius, il est hautement vraisemblable que des régions régulatrices s'étendent hors du fragment de promoteur de 1.7 kb que nous avons caractérisé. En vue de préparer la caractérisation ultérieure d'autres régions, nous avons isolé un clone PAC d'ADN génomique souris qui contient un insert de quelque 160 kb, avec 20 kb de région 5', l'exon 1, 2, 3 et 4 ainsi que l'intron 1 dont la taille se situe autour de 70-80 kb. Nous avons montré que ce clone n'est probablement pas réarrangé, ce qui reste à confirmer par des analyses complémentaires de son contenu en exons *reelin*.

La présence dans les banques de données d'une partie de la séquence promotrice du gène *reelin* humain nous a permis de comparer les deux promoteurs. Le promoteur *reelin* humain contient aussi un îlot CpG. D'une taille de 1.2kb, il couvre la région promotrice, l'exon 1 et le début de l'intron 1 et possède un contenu G+C de 75 %, dont 93 % de CpG. Le contenu en GC du promoteur humain est plus élevé que chez la souris. Ce phénomène est appelé érosion des îlots CpG, et est attribué à une intense méthylation des îlots suivie de mutation en TpG (déamination en thymines des 5-méthylcytosines). L'alignement des séquences des deux promoteurs fait apparaître la présence de 5 régions conservées. La région couvrant le site d'initiation de la transcription est extrêmement bien conservée (85 %), et des éléments régulateurs fondamentaux pour l'expression du gène *reelin* y sont probablement localisés.

Les sites potentiels de fixation des facteurs de transcription ont été identifiés par le programme MatInspector genomatix (www.genomatix.de).

Nous avons retenu les sites conservés chez l'homme et la souris et qui présentent une homologie de 100 % avec le site minimal de liaison du facteur. Dix sites de liaisons potentiels ont été décelés. Plusieurs facteurs de transcription potentiels ont une expression ubiquiste (**RFX1**, **AhR**, **SP1**, **AP2** et **AP4**). Un autre, **IRF2**, est un régulateur négatif, constitutif de l'interféron β . Enfin, deux d'entre eux, **PAX5** et **IK2** sont exprimés au niveau du système nerveux. Le premier est nécessaire pour le développement du mésencéphale, et le second est exprimé au niveau du striatum. Ces programmes de prédiction sont utiles pour suggérer certaines pistes. Cependant, seule une analyse expérimentale du promoteur permettra de déterminer les facteurs transcriptionnels physiologiquement actifs.

Comme première approche, nous avons entamé l'étude de la région promotrice de la souris par transfections transitoires. Ces expériences ont permis de définir des régions en amont du site d'initiation de transcription qui semble exercer un effet stimulant, ainsi qu'une région dont l'effet semble plutôt inhibiteur. Ces expériences devront être répétées afin d'obtenir une meilleure statistique, et comparées aux résultats obtenus avec le promoteur humain.

Une fois définies quelques régions candidates, une étude biochimique plus ciblée peut être entreprise, au moyen des deux techniques classiques de retard sur gel et de protection de la DNase. La technique de retard sur gel ou « Electrophoretic Mobility Shift Assay » (EMSA) permet de mettre en évidence la (non)-liaison d'une protéine avec une séquence génomique donnée. Un oligonucléotide contenant la séquence cible d'un facteur de transcription est incubé en présence d'un extrait cellulaire avant d'être déposé sur un gel d'électrophorèse en conditions non dénaturantes. Si la protéine est présente sous sa forme active dans l'extrait cellulaire, elle se lie à la séquence cible et la migration du complexe acides nucléiques-protéine ainsi formé est plus lente que celle de la sonde libre. La technique de protection à la DNase ou analyse par empreinte (footprinting) est intéressante pour localiser les sites spécifiques de liaisons des facteurs transcriptionnels. Ce test est basé sur le fait que la protéine, une fois liée à son site, protège l'ADN de l'activité hydrolytique de la DNase.

Dans un cadre plus général, notre laboratoire est davantage intéressé par la définition des régions génomiques qui procurent la spécificité d'expression de la *reelin* dans les neurones de la zone ventriculaire que dans l'étude détaillée de la fixation des facteurs de régulation au niveau du promoteur proximal. Cette définition nécessite d'étudier l'expression de *reelin* in vivo par les neurones mêmes, ce qui ne peut se réaliser actuellement que par deux techniques, à savoir l'électroporation in vivo et surtout la production de souris transgéniques. L'électroporation in vivo consiste à introduire un plasmide contenant un fragment de *reelin* (candidat promoteur à tester) en amont d'un rapporteur (par ex. LacZ), par électroporation de l'embryon entier ou d'un explant comme par exemple un hémisphère cérébral embryonnaire au 14^{ème} jour. Il s'agit d'une technique de plus en plus utilisée chez le poulet et qui a été récemment introduite chez la souris. Quelques essais ont été réalisés au laboratoire, mais montrent que cette méthode n'est pas facilement transposable à l'étude de l'expression dans les cellules de Cajal-Retzius. C'est pourquoi nous pensons que l'approche par souris transgéniques est actuellement l'outil de choix. Dans cette approche, des fragments de DNA de la région à tester sont micro-injectés dans des ovocytes en même temps qu'un rapporteur tel que LacZ. Les ovocytes sont transférés chez une souris porteuse et le pattern d'expression du rapporteur est étudié au stade embryonnaire intéressant. Une alternative est de cloner le rapporteur directement en aval du fragment de DNA à tester. En étudiant des embryons transgéniques produits par des régions différentes, il est théoriquement possible de définir les fragments de DNA qui sont nécessaires à l'expression de *reelin* dans les neurones de Cajal-Retzius. C'est d'ailleurs en préalable de cette approche expérimentale que nous avons isolé un clone PAC de souris qui contient très probablement toutes les régions impliquées dans cette régulation spécifique. Une première série d'expériences de transgénèse serait de micro-injecter l'insert de ce clone PAC en même temps qu'une construction contenant LacZ. Comme les fragments co-injectés ont tendance à s'intégrer dans le génome en tandem, on peut espérer observer une expression de LacZ qui récapitule l'expression du gène *reelin* endogène. Sinon, une technique récemment décrite par le groupe de Stewart (Zhang et al., 1998) permet de modifier un clone génomique chez E.Coli et d'y introduire un gène rapporteur par recombinaison homologue.

Nous pensons que l'approche par transgénèse est actuellement la plus prometteuse et devrait permettre de cibler les régions les plus importantes qui permettent l'amplification de la synthèse de *reelin* dans les neurones de la zone marginale. Si cette hypothèse est exacte, une étude détaillée de ces régions devrait permettre de mieux comprendre le développement du cortex cérébral des mammifères et peut-être contribuer à éclairer certains des mécanismes qui ont conduit à l'évolution spectaculaire du cortex depuis les amniotes-souches jusqu'aux mammifères.

Bibliographie

V. Bibliographie

Anton ES, Kreidberg JA, Rakic P (1999). Distinct functions of alpha-3 and alpha (v) integrin receptors in neuronal migration and laminar organization of the cerebral cortex. *Neuron*. **22** : 277-89

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* **215** : 403-410

Bar I, Lambert de Rouvroit C, Royaux I, Krizman DB, Dernoncourt C, Ruelle D, Beckers MC, Goffinet AM (1995). A YAC contig containing the reeler locus with preliminary characterization of candidate gene fragments. *Genomics*. **26** : 543-549

Bain G, Ray WJ, Yao M, Gottlieb DI (1994). From embryonal carcinoma cells to neurons : the P19 pathway. *BioEssays*. **16** : 343-348

Bernier B, de Bergeyck V, Lambert de Rouvroit C, Royaux I, Goffinet AM (1998). Reelin et développement cérébral : état de la question depuis le clonage du gène. *Médecine/Sciences*. **14** : 637-43

Bird A (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3** : 342-347

Bishop KM, Goudreau G, O'Leary DD (2000). Regulation of area identity in mammalian neocortex by Emx2 and Pax6. *Science*. **288** : 344-349

Bickmore WA, Bird AP (1992). Use of restriction endonucleases to detect and isolate gens from mammalian cells. *Methods of Enzymol.* **216** : 224-244

Blackwood EM, Kadonaga JT (1998). Going the distance : a current view of enhancer action. *Science*. **281** : 60-63

Buono P, de Concillis L, Olivetta E, Izzo P, Salvatore F (1993). Cis-acting elements in the promoter region of the human aldolase C gene. *FEBS Lett.* **328** : 243-249

Chae T, Kwon YT, Bronson R, dikkes p, Li E, Tsai LH (1997). Mice lacking p35, a neuronal specific activator of Cdk 5, display cortical lamination defects, seizures, and adult lethality. *Neuron.* **18** : 29-42

Clark JS, Harrison J, Paul CL, Frommer M (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Research.* **22** : 2990-2997

D'Arcangelo G, Miao GG, Chen SC, Soares HD, Morgan JI, Curran T (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374** : 719-723

D'Arcangelo G, Nakajima K, Miyata T, Ogawa M, Mikoshiba K, Curran T (1997). Reelin is a secreted glycoprotein recognized by the CR-50 monoclonal antibody. *J. Neuroscience* **17** : 23-31

D'Arcangelo G, Homayouni R, Keshvara L, Rice DS, Sheldon M, Curran T (1999). Reelin is a ligand for lipoprotein receptors. *Neuron* **24** : 471-479

Gardinner-Garden, Frommer (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196** : 261-282

Goffinet AM (1984). Events governing organization of postmigratory neurons : studies on brain development in normal and reeler mice. *Brain Research.* **7** : 261-296

Hevner RF, Shi L, LeWinter R, Rubenstein JLR (1999). Cortical phenotype of Tbr-1 mutant mice : laminar inversion and decreased expression of Reelin. *Soc. Neurosci. Abstr.* **25** : 502

Hiesberger T, Trommsdorff M, Howell BW, Goffinet AM, Mumby MC, Cooper JA, Herz J (1999). Direct binding of reelin to VLDL receptor and ApoE

receptor2 induces tyrosine phosphorylation of disabled-1 and modulates tau phosphorylation. *Neuron*. **24** : 481-489

Holliday R, Pugh JE (1975). DNA modification mechanisms and gene activity during development. *Science*. **187** : 226-232

Howell BW, Herrick TM, Cooper JA (1999). Reelin-induced tyrosine phosphorylation of disabled-1 during neuronal positioning. *Genes dev*. **13** : 643-648

Itasaki N, Bel-Vialar S, Krumlauf R (1999). Shocking developments in chick embryology : electroporation and in vivo gene expression. *Nature Cell Biology*. **1** : E204

Jones PA (1999). The DNA methylation paradox. *Trends Genet*. **15** : 34-37

Lambert de Rouvroit C, Goffinet AM (1998). A new view of early cortical development. *Biochemical Pharmacology*. **56** : 1403-1409

Lambert de Rouvroit C, Bernier B, Royaux I, de Bergeyck V, Goffinet AM (1999a). Evolutionarily conserved, alternative splicing of Reelin during brain development. *Neurol*. **156** : 229-238

Lambert de Rouvroit C, de Bergeyck V, Cortvrindt C, Bar I, Eeckout Y, Goffinet AM (1999b). Reelin, the extracellular matrix protein deficient in reeler mutant mice, is processed by a metalloproteinase. *Neurol*. **156** : 214-217

Larsen F, Gundersen G, Prydz H (1992). Choice of enzymes for mapping based on CpG islands in the human genome. *GATA*. **9** : 80-85

Mallamaci A, Mercurio S, Muzio L, Cecchi C, Pardini CL, Gruss P, Boncinelli E (2000). The lack of Emx2 causes impairment of Reelin signaling and defects of neuronal migration in the developing cerebral cortex. *J. Neurosci*. **20** : 1109-18

Ogbourne S, Antalís TM (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* **331** : 1-14

Pedersen AG, Baldi P, Chauvin Y, Brunak S (1999). The biology of eukariotic promoter prediction. *Computers & Chemistry.* **23** : 191-207

Purves D, Augustine GJ, Fitzpatrick D, Katz LC, Lamantia AS, McNamara JO (1999). Neurosciences. *Neurosciences & cognition.* 377-394

Rakic P (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145** : 61-84

Reiner O, Albrecht U, Gordon M, Chianese KA, Wong C, Gal-Gerber O, Sapir T, Siracusa LD, Buchberg AM, Caskey CT, Eichele G (1995). Lissencephaly gene (LIS1) expression in the CNS suggests a role in neuronal migration. *J. Neuroscience.* **15** : 3730-3738

Royaux I, Lambert de Rouvroit C, D'Arcangelo G, Demirov D, Goffinet AM (1997). Genomic organization of the mouse reelin gene. *Genomics.* **46** : 240-250

Sanes JR (1989). Extracellular matrix molecules that influence neural development. *Annu. Rev. Neurosci.* **12** : 491-516

Sciffmann SN, Bernier B, Goffinet AM (1997). Reelin mRNA expression during mouse brain development. *European J. of Neuroscience.* **9** : 1055-1071

Senzaki K, Ogawa M and Yagi T (1999). Proteins of the CNR family are multiple receptors for reelin. *Cell.* **99** : 635-647

Takaoka AS, Yamada T, Gotoh M, Kanai Y, Imai K, Hirohashi S (1998). Cloning and characterization of the human β 4-integrin gene promoter and enhancers. *J. of Biol. Chemistry.* **273** : 848-855

Tsai LH, Takahashi T, caviness VS, Harlow E (1993). Activity and expression pattern of cyclin-dependent kinase 5 in the embryonic mouse nervous system. *Development*. **119** : 1029-1040

Trommsdorff M, Gotthardt M, Hiesberger T, Shelton J, Stockinger W, Nimpf J, Hammer RE, Richardson JA, Herz J (1999). Reeler/Disabled-like disruption of neuronal migration in knockout mice lacking the VLDL receptor and ApoE receptor-2. *Cell*. **97** : 689-701

Vincent A, Heitz D, Rousseau F, Devys D, saccone S, Abderrahim H, Le Paslier D, Cohen D, tonolio D, Della Valle G (1991). Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. *Science*. **251** : 1236-9

Wingender E, Kel AE, Kel OV, Karas H, heinemeyer T, Dietze P, Knüppel R, Romaschenko AG and Kolchanov NA (1997). TRANSFAC, TRRD and COMPEL : towards a federated database system on transcriptional regulation. *Nucleic Acids Res*. **25** : 265-268

Walsh CP and Bestor TH (1999). Cytosine methylation and mammalian development. *Genes & Development*. **13** : 26-34

Walsh CA, Goffinet AM (2000). Potential mechanisms of mutations that affect neuronal migration in man and mouse. *Genetics & Development*. **10** : 270-274

Yang A, Walker N, Bronson R, Kaghad M, Oosterwegel M, bonnin J, Vagner C, bonnet H, Dikkes P, Sharpe, McKeon F, Caput D (2000). P73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature*. **404** : 99-103

Zhang Y, Buchholz F, Muirers JPP, Stewart AF (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature genetics*. **20** : 123-128

Annexes

Annexe 1 : Liste des amorces spécifiques des vecteurs utilisés

Amorce	Vecteur	Séquence amorce (5'-3')	Taille	Tm (°C)
BSL	pBluescript	ATT AAC CCT CAC TAA AGG GA	20	56
BSR	pBluescript	AAT ACG ACT CAC TAT AGG GC	20	58
pPACL	pPAC4	GCG GCC GCT AAT ACG ACT CAC TAT	24	74
pPACR	pPAC4	GGC CGT CGA CAT TTA GGT GAC ACT	24	74

Annexe 2 : Liste des amorces spécifiques du gène reelin

Amorce		Position ADNc	Séquence amorce (5'-3')	Taille	Tm (°C)
P3	->	-1692 à -1672	TGT AGA CAG TGG CAG GCC AG	20	64
Cplt P3	<-	-1672 à -1692	CTG GCC TGC CAC TGT CTA CA	20	64
P5	<-	-1532 à -1553	GCA GGC TGC ACA GGC TTG TAG	21	68
P4	<-	-1224 à -1245	ATG CAT TTC TCC AGG CTC AGC	21	64
P1	->	-766 à -745	TCC GGA CTG ACC TTT CCC GAGC	22	68
Jp1R	<-	-746 à -766	AGC CCT TTC CAG TCA GGC CT	20	64
Jp2R	<-	-627 à -647	TCT ACA CCG GCT GCA ACC CT	20	64
UPF	<-	-311 à -332	CCT CCG CAC CTT CTT AAA GCC	21	66
Cplt UPF	->	-332 à -311	GGC TTT AAG AAG GTG CGG AGG	21	66
P2	<-	-35 à -55	GCT CTC TCA TCC ACT TTC GG	20	62
Signal1	->	49 à 68	CTG CTG CTG GCG ACG CTG AG	20	68
RNASPL1	<-	188 à 168	GCA ATG TGC AGG GAA ATG AGC	21	64
4.5	->	256 à 273	GAT GGC TTG CTG GTG ACG	18	58
IRO7	<-	303 à 284	AGA CTG GAT GCT TGT CGA G	20	62
Cpl 4.9	->	376 à 395	ATG TGC AGT GTG GTG GCC TC	20	64
4.4	<-	509 à 492	AAA ATC ACC TGG CCC CTA	18	54
Rlex1	<-	183 à 207	TAGTAGGTGGGGTTGCCCCGCAATG	24	76
Rlex2	->	268 à 292	GTGACGGGACTCTATACCTCGACA	24	74

Annexe 3 : Séquence du promoteur reelin souris avec traduction

Les chiffres indiquent la numérotation en nucléotides, 1 étant le premier nucléotide du codon d'initiation. Les nucléotides en rouge indiquent les sites d'initiation de la transcription.

-1691		gggctgtagac	-1681
-1680	agtggcaggccaggcatctttagagtcagccttcgtcttacttggggcaagctattctgatttctccctgtgaaatggg		-1601
-1600	ggaagttttgcttcttcgggtggtggtgtaggattacattaggctaatagatctacaagcctgtgcagcctgcatggcac		-1521
-1520	tttgctctagtgtaatattagtcgcgcgtgatgaatatgtgtgctgttgaaactttgggaggattggaagtaatggagacc		-1441
-1440	aacaggcagtttaggtccttcctatctgctgaaactgtgaatgtggcagaaaaaaaaaatcttttgcaatgaaattaccac		-1361
-1360	agaggctctccctttggatcttgccctctgacgcccttctccaaaactcagcaagactaccatttctgacattacaagtt		-1281
-1280	tcctgggtttctttacaagcaaggcttattttggcttggtgctgagcctggagaaatgcattaaaacctggcagctctgtt		-1201
-1200	cttcccgctctctgggtgcctcggggctcctgccagggtgctcgggttaaaaaactgcttctgggatggtttcacttgc		-1121
-1120	ccgcagtcctgcgtgagatagctcattaacgcgggtttcatttgtgctgccagccatccttgcctgccttccctcccc		-1041
-1040	accccgagtgacagtggtcatgcacatgtagaagtaataaaactcggacctaataatttacttaccctccatctccgtccg		-961
-960	aggaatcccttagcaacgcgtaataacagagtcgggcgaagctcgcctgtcgggtccctccaccacaatccccacgggcc		-881
-880	gagaagcccgaacatctggccccgcaccactccaggtggcaatcgcacgcgcacccaatctccctttcagcctggctgcc		-801
-800	actgaagaccggggcgggggcagcgacggggcgagcgctgactggaaagggtcgcgcggcgggggcaccttcgcgg		-721
-720	ttggggccactgacgggaggtccccagacaagcgtgcccggtggggactggggggcgggcgcccccgagggggcaggggtt		-641
-640	gcagccggtgtagacaaagaacagcgcggtccccgcgccttcggtttctcgcgggtcctcgcagaacgggcagccgtga		-561
-560	gccgggggttcgggttccccggggcagcccgctcgggtcttgccgctttcgttcgcgcagcgccgagcagccctgcccttc		-481
-480	cgggcgagcgcccggtggggcgggagggcgcgctgtgttcgggggcacagcccgaggtgcgggcaagcccgagggggcg		-401
-400	ggacggggcgcgggggcgggggcgcggtgggacacgtgtggcgggcgggggagggggcgggcgcccggggctttaagaa		-321
-320	ggtgcggagggggcgggcactttcccaagcccgccgagggggcgtcgcgtgcacaccggcgggcgggcgctcggagggc		-241
-240	gacgacgcgtctcggcgcccgcgggcccggttcccccgcgctctcgtccggcgggcccaaagtaacttcgggagcctc		-161
-160	ggtctcccgctaacttcccccgcggggtcgggtgcccgagcccgctcgggtcagcccgcccgcggtcgccttccccg		-81
-80	cacgcgggtcctccgtgccggtgcctccgaaagtggatgagagagcgcgcgggggcgcgcgggcgacggagcgcgggc		-1
1	ATG GAG CGC GGC TGC TGG GCG CCG CGG GCT CTC GTC CTG GCC GTG CTG CTG CTG CTG GCG		60
1	M E R G C W A P R A L V L A V L L L L A		20
61	ACG CTG AGG GCG CGC GCG GCC ACC GGC TAC TAC CCG CGC TTC TCG CCT TTC TTT TTC CTG		120
21	T L R A R A A T G Y Y P R F S P F F F L		40
121	TGC ACC CAC CAC GGG GAG CTG GAA GGG GAT GGG GAG CAG GGC GAG GTG CTC ATT TCC CTG		180
41	C T H H G E L E G D G E Q G E V L I S L		60

Annexe 4 : Séquence du promoteur reelin humain avec traduction

Les chiffres indiquent la numérotation en nucléotides, 1 étant le premier nucléotide du codon d'initiation. Les nucléotides en rouge indiquent les sites d'initiation de la transcription.

```

-1760tacgtctagccccacagcctgcaccccgagagctacagtgtagcaattgttaccacatcatatatatatatgcacat-1681
-1680acatatatgtatacactgtatattatttatatggcagttaagcctctggggatthttgaagtaagggaacaagaagccag-1601
-1600ctagaaatgggcccctttctacctgtcattcgggtctcagtaagaggaagaaaaacaattttcttcccttttgaacgaagtga-1521
-1520ttacagaggtctccccttggaccctgattttcaccgcccctcgccccacaaatgcaggcacaaggcactatcccctccaat-1441
-1440tttataatctactcattttctgtaogtgcgggtttatctggttggtggcaagccaagggtgggtggggagaggcattaaa-1361
-1360acctggcaaccccgctcccccattctcgggatgcgcagggtcctcgccagccgtcctcggttaaaaaagtgcttatggat-1281
-1280tatttcacttcctgcagttcatgtgcagccgctcattaccacccgcttcaatcacgctgccagcatctcccctgaacg-1201
-1200tttttccctcccccgatttcagggccattgggtcataaccgacatttaaaggaaataaacactgatttaaaaaattacaac-1121
-1120accatctctgcctgaagagcccagaagcaatgaataacagaacagtcggggaagagagcctgtcggtatcttagctccg-1041
-1040ccagcccttctgccttacaattccccacacccgagcagcacaaatcaccattttccaggccaccactcgagtctgccaatt -961
-960 ccactttcgcttaattctcattttaaaatcctctacagatgaggctctgtcgtgcgcgagggggccgggcccgcggcg -881
-880 ggagggaacagggggcctgggtgggaagggagcgcgagggcccttagcggttcgcggcctcgggcggtattcccagaacgag -801
-800 tgcggggccgcccggactggcagactgcgagccccgagggggcgcggggacgcagccggtcagacaaaagaaaccggcgccag -721
-720 cggccgtccctgcgcgccctctccttccctcaogcatcctcccaggaaaaacagggcacactgacggccaaaggggctgg -641
-640 ccttcccttcagacagctggacgggaggtgccctctgcggggctttgacgtccctcgagaagagtcgcgggctcagcg -561
-560 gtcctcgacagcgtcccggtcccggtcccccggggcgccccctccctgtcctcccggtgcgaaccggggcgctggccggg -481
-480 gactccgggggacgcgtgcgcccctcgccgcgcgaggtgccgcgagccagcccgagagggggggggggcgggggcgggc -401
-400 gcgccccgggggggagcggccgggacacgtgtggggcggggggggggacgcggcgccccggggctttaagaaggtgtgg -321
-320 agcggggcgggcgctttcccaggcctggccgagggggcgtcgcgcagagggcgggcgggcgccagggcgggcagacgacg -241
-240 cgtctcgcggcgcccgagccccggtcccgctcccgcgccccaaagtaactttgggagccgctctcccgcggaactt -161
-160 ctcgggggggggccgctcccaggcccgctcccagccccgttcgcgtcccgctccgcttcttctcgccttctctccgcgtg -81
-80 gtcctccgtcccggtctccaaaactgaatgagcgcgagcgggcgctagggcgcgggcgggcgggcgggcgggcgggc -1

1  ATG GAG CGC AGT GGC TGG GCC CGG CAG ACT TTC CTC CTA GCG CTG TTG CTG GGG GCG ACG      60
1  M  E  R  S  G  W  A  R  Q  T  F  L  L  A  L  L  L  G  A  T      20

61  CTG AGG GCG CGC GCG GCT GGC TAT TAC CCC CGC TTT TCG CCC TTC TTT TTC CTG TGC      120
21  L  R  A  R  A  A  A  G  Y  Y  P  R  F  S  P  F  F  F  L  C      40

121 ACC CAC CAC GGG GAG CTG GAA GGG GAT GGG GAG CAG GGC GAG GTG CTC ATT TCC CTG CAT      180
41  T  H  H  G  E  L  E  G  D  G  E  Q  G  E  V  L  I  S  L  H      60

181 ATT GCG GGC AAC CCC ACC TAC TAC GTT CCG GGA CAA GAA TAC CAT G      240
61  I  A  G  N  P  T  Y  Y  V  P  G  Q  E  Y  H      75

```